

# INEXACT ALTERNATING DIRECTION MULTIPLIER METHODS FOR SEPARABLE CONVEX OPTIMIZATION \*

WILLIAM W. HAGER<sup>†</sup> AND HONGCHAO ZHANG<sup>‡</sup>

**Abstract.** Inexact alternating direction multiplier methods (ADMMs) are developed for solving general separable convex optimization problems with a linear constraint and with an objective that is the sum of smooth and nonsmooth terms. The approach involves linearized subproblems, a back substitution step, and either gradient or accelerated gradient techniques. Global convergence is established. The methods are particularly useful when the ADMM subproblems do not have closed form solution or when the solution of the subproblems is expensive. Numerical experiments based on image reconstruction problems show the effectiveness of the proposed methods.

**Key words.** Separable convex optimization, Alternating direction method of multipliers, ADMM, Multiple blocks, Inexact solve, Global convergence

**AMS subject classifications.** 90C06, 90C25, 65Y20

**1. Introduction.** We consider a convex separable linearly constrained optimization problem

$$(1.1) \quad \min \Phi(\mathbf{x}) \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}$$

where  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  and  $\mathbf{A}$  is  $N$  by  $n$ . By a separable convex problem, we mean that the objective function is a sum of  $m$  independent parts, and the matrix is partitioned compatibly as in

$$(1.2) \quad \Phi(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x}_i) + h_i(\mathbf{x}_i) \quad \text{and} \quad \mathbf{Ax} = \sum_{i=1}^m \mathbf{A}_i \mathbf{x}_i.$$

Here  $f_i$  is convex and Lipschitz continuously differentiable,  $h_i$  is a proper closed convex function (possibly nonsmooth),  $\mathbf{A}_i$  is  $N$  by  $n_i$  with  $\sum_{i=1}^m n_i = n$ , and the columns of  $\mathbf{A}_i$  are linearly independent for  $i \geq 2$ . Constraints of the form  $\mathbf{x}_i \in \mathcal{X}_i$ , where  $\mathcal{X}_i$  is a closed convex set, can be incorporated in the optimization problem by setting  $h_i(\mathbf{x}_i) = \infty$  when  $\mathbf{x}_i \notin \mathcal{X}_i$ . The problem (1.1)–(1.2) has attracted extensive research due to its importance in areas such as image processing, statistical learning and compressed sensing. See the recent survey [3] and its references.

Let  $\mathcal{L}$  be the Lagrangian given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \Phi(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{Ax} - \mathbf{b} \rangle,$$

where  $\boldsymbol{\lambda}$  is the Lagrange multiplier for the linear constraint and  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product. It is assumed that there exists a solution  $\mathbf{x}^*$  to (1.1)–(1.2) and an associated Lagrange multiplier  $\boldsymbol{\lambda}^* \in \mathbb{R}^N$  such that  $\mathcal{L}(\cdot, \boldsymbol{\lambda}^*)$  attains a minimum at  $\mathbf{x}^*$ , or equivalently, the following first-order optimality conditions hold:  $\mathbf{Ax}^* = \mathbf{b}$  and for  $i = 1, 2, \dots, m$  and for all  $\mathbf{u} \in \mathbb{R}^{n_i}$ , we have

$$(1.3) \quad \langle \nabla f_i(\mathbf{x}_i^*) + \mathbf{A}_i^T \boldsymbol{\lambda}^*, \mathbf{u} - \mathbf{x}_i^* \rangle + h_i(\mathbf{u}) \geq h_i(\mathbf{x}_i^*),$$

\*April 8, 2016. The authors gratefully acknowledge support by the National Science Foundation under grants 1522629 and 1522654, and by the Office of Naval Research under grant N00014-15-1-2048.

<sup>†</sup>hager@math.ufl.edu, <http://people.clas.ufl.edu/hager/>, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 294-2308. Fax (352) 392-8357.

<sup>‡</sup>hozhang@math.lsu.edu, <http://www.math.lsu.edu/~hozhang>, Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918. Phone (225) 578-1982. Fax (225) 578-4276.

where  $\nabla$  denotes the gradient.

A popular strategy for solving (1.1)–(1.2) is the alternating direction multiplier method (ADMM) [15, 16] given by

$$\begin{cases} \mathbf{x}_i^{k+1} &= \arg \min_{\mathbf{x}_i \in \mathbb{R}^{n_i}} L(\mathbf{x}_1^{k+1}, \dots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i, \mathbf{x}_{i+1}^k, \dots, \mathbf{x}_m^k, \boldsymbol{\lambda}^k), \quad i = 1, \dots, m, \\ \boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}), \end{cases} \quad (1.4)$$

where  $L$ , the augmented Lagrangian, is defined by

$$(1.5) \quad L(\mathbf{x}, \boldsymbol{\lambda}) = \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2.$$

Here  $\rho > 0$  is the penalty parameter. Early ADMMs only consider problem (1.1)–(1.2) with  $m = 2$  corresponding to a 2-block structure. In this case, the global convergence and complexity can be found in [12, 25]. When  $m \geq 3$  the ADMM strategy (1.4), a natural extension of the 2-block ADMM, is not necessarily convergent [5], although its practical efficiency has been observed in many recent applications [34, 35].

Recently, much research has focused on modifications to ADMM to ensure convergence when  $m \geq 3$ . References include [4, 6, 7, 11, 17, 21, 23, 24, 29, 30, 31]. One approach [4, 7, 11, 21] assumes  $m - 2$  of the functions in the objective are strongly convex and the penalty parameter is sufficiently small. Linear convergence results under additional conditions are obtained in [31]. Analysis of a randomly permuted ADMM which allows for nonseparated variables is given in [6]. Another approach, first developed in [23, 24], involves a back substitution step to complement the ADMM forward substitution step. The algorithms developed in our paper utilize this back substitution step.

The dominant computation in an iteration of ADMM is the solution of the subproblems in (1.4). The efficiency depends on our ability to solve these subproblems inexactly while still maintaining the global convergence, especially when no closed formula exists for the subproblems [36]. One line of research is to solve the subproblems to an accuracy based on an absolute summable error criterion [9, 12, 18]. In [28], the authors combine an adaptive error criterion with the absolute summable error criterion for 2-block ADMM with logarithmic-quadratic proximal regularization and further correction steps to modify the solutions generated from the ADMM subproblems. In [14], the authors develop a 2-block ADMM with a relative error stopping condition for the subproblems, motivated by [13], based on the total subgradient error. Another line of research is to add proximal terms to make the subproblems strongly convex [8, 22] and relatively easy to solve. However, this approach often requires accurate solution of the proximal subproblems. When  $m = 1$ , ADMM reduces to the standard augmented Lagrangian method (ALM), for which practical relative error criteria for solving the subproblems have been developed and encouraging numerical results have been obtained [13, 33]. In this paper, motivated by our recent work on variable stepsize Bregman operator splitting methods (BOSVS), by recent complexity results for gradient and accelerated methods for convex optimization, and by the adaptive relative error strategy used in ALM, we develop new *inexact* approaches for solving the ADMM subproblems. To the best of our knowledge, these are the first ADMMs for solving the general separable convex optimization problem (1.1)–(1.2) based on an adaptive accuracy condition that does not employ an absolute summable error criterion and that guarantees global convergence, even when  $m \geq 3$ .

To guarantee global convergence, a block Gaussian backward substitution strategy is used to make corrections to the approximate subproblem solutions. In the special

case  $m = 2$ , the method will reduce to a 2-block ADMM without back substitution. This idea of using block Gaussian back substitution was first proposed in [23, 24]. The method in this earlier work requires the exact solution of the subproblems to obtain global convergence, while our new approach allows an inexact solution. More recently, a linearly convergent ADMM was developed in [26]. This algorithm linearizes the subproblems to achieve an inexact solution, and requires that the functions  $f_i$  and  $h_i$  in the objective function satisfy certain “local error bound” conditions. In addition, to ensure linear convergence, the stepsize  $\alpha_k$  in (1.4) must be sufficiently small, which could significantly deteriorate the practical performance.

In this paper, we focus on problems where the minimization of the dual function over one or more of the primal variable  $\mathbf{x}_i$  is nontrivial, and the accuracy of an inexact minimizer needs to be taken into account. On the other hand, when these minimizations are simple enough, it is practical to minimize the Lagrangian over  $\mathbf{x}$  and compute the dual function. This leads to a possibly nonsmooth dual function which can be approached through smoothing techniques as in [27], or through active set techniques as in [20].

Our paper is organized as follows. In the Section 3, we first generalize the BOSVS algorithm [10, 19] to handle multiple blocks. The original BOSVS algorithm was tailored to the two block case, but used an adaptive stepsize when solving the subproblem, and consequently, it achieved much better overall efficiency when compared to the Bregman operator splitting (BOS) type algorithms based on a fixed smaller stepsize. In Sections 4 and 5, more adaptive stopping criteria for the subproblems are proposed. The adaptive criteria for bounding the accuracy in the ADMM subproblems are based on both the current and accumulated iteration change in the subproblem. These novel stopping criteria are motivated by the complexity analysis of gradient methods for convex optimization, and by the relative accuracy strategy often used in an inexact augmented Lagrangian method for nonlinear programming. Although our analysis is carried out with vector variables, these results could be extended to matrix variables which could have more potential applications.

**1.1. Notation.** The set of solution/multiplier pairs for (1.1) is denoted  $\mathcal{W}^*$ , while  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$  is a generic solution/multiplier pair. For  $\mathbf{x}$  and  $\mathbf{y} \in \mathbb{R}^n$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$  is the standard inner product, where the superscript  $\top$  denotes transpose. The Euclidean norm, denoted  $\|\cdot\|$ , is defined by  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$  and  $\|\mathbf{x}\|_{\mathbf{G}} = \sqrt{\mathbf{x}^\top \mathbf{G} \mathbf{x}}$  for a positive definite matrix  $\mathbf{G}$ .  $\mathbb{R}^+$  denotes the set of nonnegative real numbers, while  $\mathbb{R}^{++}$  denotes the set of positive real numbers. For a differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\nabla f(\mathbf{x})$  is the gradient of  $f$  at  $\mathbf{x}$ , a column vector. More generally,  $\partial f(\mathbf{x})$  denotes the subdifferential at  $\mathbf{x}$ . If  $\mathbf{x}$  is a vector, then  $\mathbf{x}_+$  denotes the subvector obtained by dropping the first block of variables from  $\mathbf{x}$ . Thus if  $\mathbf{x} \in \mathbb{R}^n$  with  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  for  $i \in [1, m]$ , then  $\mathbf{x}_+ = (\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_m)^\top$ .

**2. Algorithm Structure.** Three related inexact ADMMs are developed called generalized, multistep, and accelerated BOSVS. They differ in the details of the formula for the new iterate  $\mathbf{x}^{k+1}$ , but the overall structure of the algorithms is the same. Both multistep and accelerated BOSVS typically represent a more exact ADMM iteration when compared to generalized BOSVS, while the accelerated BOSVS subiterations often converge more rapidly than those of multistep BOSVS. The common elements of these three algorithms appear in Algorithm 2.1.

The algorithms generate three sequences  $\mathbf{x}^k$ ,  $\mathbf{y}^k$ , and  $\mathbf{z}^k$ . In Step 1 of Algorithm 2.1 there may be more than one ADMM subiteration, as determined by an adaptive stopping criterion. The iterate  $\mathbf{x}^{k+1}$  is the final iterate generated in the

---

**Parameters:**  $\rho, \sigma, \delta_{\min}, \theta_1, \theta_2, \theta_3 \in \mathbb{R}^{++}$ ,  $\delta_{\min} < \delta_{\max}$ ,  $\alpha \in (0, 1)$ ,  $\sigma < 1 < \tau \leq \eta$ .

**Starting guess:**  $\mathbf{x}^1$  and  $\boldsymbol{\lambda}^1$ .

**Initialize:**  $\mathbf{y}^1 = \mathbf{x}^1$ ,  $k = 1$ ,  $\delta_{\min, i} = \delta_{\min}$  and  $\Gamma_i^0 = 0$ ,  $1 \leq i \leq m$ ,  $e^0 = \infty$

**Step 1:** For  $i = 1, \dots, m$

Generate  $\mathbf{x}_i^{k+1}$  and  $\mathbf{z}_i^k$ , estimate  $r_i^k \approx \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2$ .

End

**Step 2:** If  $e^k = \theta_1 \|\mathbf{z}_+^k - \mathbf{y}_+^k\| + \theta_2 \|\mathbf{A}\mathbf{z}^k - \mathbf{b}\| + \theta_3 \sqrt{\sum_{i=1}^m r_i^k}$  is sufficiently small, terminate.

**Step 3:** Set  $\mathbf{y}_1^{k+1} = \mathbf{z}_1^k$ ,  $\mathbf{y}_+^{k+1} = \mathbf{y}_+^k + \alpha \mathbf{M}^{-\top} \mathbf{H}(\mathbf{z}_+^k - \mathbf{y}_+^k)$ ,  
 $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha \rho (\mathbf{A}\mathbf{z}^k - \mathbf{b})$ ,  $k := k + 1$ , and go to Step 1.

---

ALG. 2.1. Our ADMM structure.

ADMM (forward substitution) subproblems,  $\mathbf{y}^k$  is generated by the back substitution process in Step 3, and  $\mathbf{z}^k$  is an average of the iterates in the ADMM subproblems of Step 1. In generalized BOSVS,  $\mathbf{z}^k = \mathbf{x}^{k+1}$  since there is only one ADMM subiteration, while multistep and accelerated BOSVS typically perform more than one subiteration and  $\mathbf{z}^k$  is obtained by a nontrivial averaging process. The matrix  $\mathbf{M}$  in Step 3 is the  $m - 1$  by  $m - 1$  block lower triangular matrix defined by

$$(2.1) \quad \mathbf{M}_{ij} = \begin{cases} \mathbf{A}_{i+1}^\top \mathbf{A}_{j+1} & \text{if } 1 \leq j \leq i < m, \\ \mathbf{0} & \text{if } 1 \leq i < j < m. \end{cases}$$

The matrix  $\mathbf{H}$  is the  $m - 1$  by  $m - 1$  block diagonal matrix whose diagonal blocks match those of  $\mathbf{M}$ . The matrices  $\mathbf{M}$  and  $\mathbf{H}$  are invertible since the columns of  $\mathbf{A}_i$  are linearly independent for  $i \geq 2$ , which implies that  $\mathbf{A}_i^\top \mathbf{A}_i$  is invertible for  $i \geq 2$ .

**3. Generalized BOSVS.** Our first algorithm is a generalization of the BOSVS algorithm developed in [10, 19] for a two-block optimization problem. Let  $\Phi_i^k : \mathbb{R}^{n_i} \times \mathbb{R}^{n_i} \times \mathbb{R} \rightarrow \mathbb{R}$  be defined by

$$\Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) = f_i(\mathbf{v}) + \langle \nabla f_i(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle + \frac{\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2 + h_i(\mathbf{u}) + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2,$$

where

$$(3.1) \quad \mathbf{b}_i^k = \mathbf{b} - \sum_{j < i} \mathbf{A}_j \mathbf{z}_j^k - \sum_{j > i} \mathbf{A}_j \mathbf{y}_j^k.$$

The function  $\Phi_i^k$  corresponds to the part of the augmented Lagrangian associated with the  $i$ -th component of  $\mathbf{x}$ , but with the smooth term  $f_i$  linearized around  $\mathbf{v}$  and with a proximal term added to the objective. Algorithm 3.1 which follows is the Step 1 inner loop of Algorithm 2.1 for generalized BOSVS. Throughout the paper, the generalized BOSVS algorithm refers to Algorithm 2.1 with the Step 1 inner loop given by Algorithm 3.1.

Although  $\mathbf{y}^k$  does not appear explicitly in Algorithm 3.1, it is hidden inside the  $\mathbf{b}_i^k$  term of  $\Phi_i^k$ . The iterate  $\mathbf{x}_i^{k+1}$  is obtained by minimizing the  $\Phi_i^k$  function and checking the line search condition of Step 1b. In Step 1a, mid denotes median and the initial stepsize  $\delta_{i,0}^k$  of Step 1a is a safeguarded version of the Barzilai-Borwein formula [2].

**Inner loop of Step 1:**

- 
- 1a. For  $k > 1$ ,  $\delta_{i,0}^k = \text{mid} \{ \delta_{\min,i}, s^{BB}, \delta_{\max} \}$  and  
 $s^{BB} = \langle \nabla f_i(\mathbf{x}_i^k) - \nabla f_i(\mathbf{x}_i^{k-1}), \mathbf{x}_i^k - \mathbf{x}_i^{k-1} \rangle / \|\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\|^2$ .  
 For  $k = 1$ ,  $\delta_{i,0}^k$  can be any scalar in  $[\delta_{\min}, \delta_{\max}]$ .
- 1b. Set  $\delta_i^k = \eta^j \delta_{i,0}^k$ , where  $j \geq 0$  is the smallest integer such that  
 $f_i(\mathbf{x}_i^k) + \langle \nabla f_i(\mathbf{x}_i^k), \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \rangle + \frac{(1-\sigma)\delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \geq f_i(\mathbf{x}_i^{k+1})$ ,  
 where  $\mathbf{x}_i^{k+1} = \mathbf{z}_i^k = \arg \min \{ \Phi_i^k(\mathbf{u}, \mathbf{x}_i^k, \delta_i^k) : \mathbf{u} \in \mathbb{R}^{n_i} \}$ .
- 1c. Set  $r_i^k = (1/\delta_i^k) \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2$ .
- 1d. If  $k > 1$  and  $\delta_i^k > \max\{\delta_i^{k-1}, \delta_{\min,i}\}$ , then  $\delta_{\min,i} := \tau \delta_{\min,i}$ .
- 

ALG. 3.1. Inner loop in Step 1 of Algorithm 2.1 for the generalized BOSVS scheme.

Let  $\zeta_i$  denote the Lipschitz constant for  $\nabla f_i$ . By a Taylor expansion of  $f_i$  around  $\mathbf{x}_i^k$ , we see that the line search condition of Step 1b is satisfied whenever  $(1-\sigma)\delta_i^k \geq \zeta_i$ , or equivalently, when

$$(3.2) \quad \delta_i^k \geq \zeta_i / (1 - \sigma).$$

Since  $\eta > 1$ ,  $\delta_i^k$  increases as  $j$  increases, and consequently, (3.2) holds for  $j$  sufficiently large. Hence, if  $j > 0$  at the termination of the line search, we have  $\delta_i^k \leq \eta \zeta_i / (1 - \sigma)$ . If the line search terminates for  $j = 0$ , then  $\delta_i^k \leq \delta_{\max}$ . In summary, we have

$$(3.3) \quad \delta_{\min} \leq \delta_i^k \leq \max\{\eta \zeta_i / (1 - \sigma), \delta_{\max}\} \quad \text{for all } k.$$

Since  $s^{BB} \leq \zeta_i$  in Step 1a, it follows that  $\delta_{i,0}^k = \delta_{\min,i}$  whenever  $\delta_{\min,i} \geq \zeta_i$ . In Step 1d,  $\delta_{\min,i}$  is increased by the factor  $\tau$  whenever  $\delta_i^k > \delta_i^{k-1}$ . Hence, after a finite number of iterations where  $\delta_i^k > \delta_i^{k-1}$ , we have  $\delta_{\min,i} \geq \zeta_i / (1 - \sigma)$ , which implies that the line search terminates at  $j = 0$  with  $\delta_i^k = \delta_{i,0}^k = \delta_{\min,i}$ . We conclude that

$$(3.4) \quad \delta_i^k \leq \delta_i^{k-1} \quad \text{for } k \text{ sufficiently large,}$$

where  $\delta_i^k$  denotes the final accepted value in Step 1b. Note that the inequality in (3.4) cannot be replaced by equality since the number of iterations where  $\delta_i^k > \delta_i^{k-1}$  may not be enough to yield  $\delta_{\min,i} \geq \zeta_i / (1 - \sigma)$ .

In the BOS algorithm, the line search is essentially eliminated by taking  $\delta_i^k$  larger than the Lipschitz constant  $\zeta_i$ . Taking  $\delta_i^k$  large, however, causes  $\|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|$  to be small due to the proximal term in the objective  $\Phi_i^k(\cdot, \mathbf{x}_i^k, \delta_i^k)$  associated with  $\mathbf{x}_i^{k+1}$ . These small steps lead to slower convergence than what is achieved with BOSVS where  $\delta_i^k$  is adjusted by the line search criterion in order to achieve a small, but acceptable, choice for  $\delta_i^k$ .

In our analysis of generalized BOSVS, we first observe that when  $e^k = 0$ , we have reached a solution of (1.1)–(1.2).

LEMMA 3.1. *If  $e^k = 0$  in the generalized BOSVS algorithm, then  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{y}^k$  solves (1.1)–(1.2) and  $(\mathbf{x}^k, \boldsymbol{\lambda}^k) \in \mathcal{W}^*$ .*

*Proof.* Let  $\mathbf{x}^*$  denote  $\mathbf{x}^k$ . If  $e^k = 0$ , then  $r_i = 0$  for each  $i$ , and by Step 1c of generalized BOSVS,  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{x}^*$ . For generalized BOSVS, we set  $\mathbf{z}^k = \mathbf{x}^{k+1}$  in Step 1b. Since  $\mathbf{x}^{k+1} = \mathbf{x}^k$ , it follows that  $\mathbf{z}^k = \mathbf{x}^*$ . The identity  $\mathbf{z}^k = \mathbf{x}^{k+1}$  also

implies that  $\mathbf{z}^{k-1} = \mathbf{x}^k = \mathbf{x}^*$ . In Step 3 of Algorithm 2.1,  $\mathbf{y}_1^k = \mathbf{z}_1^{k-1} = \mathbf{x}_1^*$ . Since  $e^k = 0$ , Step 2 of Algorithm 2.1 implies that  $\mathbf{y}_+^k = \mathbf{z}_+^k = \mathbf{x}_+^*$ . Hence,  $\mathbf{y}^k = \mathbf{x}^* = \mathbf{z}^k$ . Since  $e^k = 0$ , it also follows from Step 2 that  $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ . Consequently, we have

$$(3.5) \quad \mathbf{b}_i^k = \mathbf{b} - \sum_{j < i} \mathbf{A}_j \mathbf{z}_j^k - \sum_{j > i} \mathbf{A}_j \mathbf{y}_j^k = \mathbf{b} - \sum_{j < i} \mathbf{A}_j \mathbf{x}_j^* - \sum_{j > i} \mathbf{A}_j \mathbf{x}_j^* = \mathbf{A}_i \mathbf{x}_i^*.$$

By Step 1b,  $\mathbf{x}_i^{k+1}$  is the minimizer of  $\Phi_i^k(\cdot, \mathbf{x}_i^k, \delta_i^k)$ . Since  $\mathbf{x}_i^{k+1} = \mathbf{x}_i^k = \mathbf{x}_i^*$ , it follows that  $\mathbf{x}_i^*$  is the minimizer of  $\Phi_i^k(\cdot, \mathbf{x}_i^*, \delta_i^k)$ . After taking into account (3.5), the first-order optimality condition associated with the minimizer  $\mathbf{x}_i^*$  of  $\Phi_i^k(\cdot, \mathbf{x}_i^*, \delta_i^k)$  is exactly the same as (1.3), but with  $\boldsymbol{\lambda}^*$  replaced  $\boldsymbol{\lambda}^k$ . Hence,  $(\mathbf{x}^*, \boldsymbol{\lambda}^k) \in \mathcal{W}^*$ .  $\square$

Two lemmas are needed for the convergence of the generalized BOSVS algorithm.

LEMMA 3.2. *Suppose that  $\mathbf{u}$  and  $\mathbf{v} \in \mathbb{R}^{n_i}$  satisfy*

$$(3.6) \quad f_i(\mathbf{v}) + \langle \nabla f_i(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle + \frac{(1-\sigma)\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2 \geq f_i(\mathbf{u})$$

for some  $\sigma \in [0, 1)$  and  $\delta > 0$ , where  $\mathbf{u}$  minimizes  $\Phi_i^k(\cdot, \mathbf{v}, \delta)$ . Then, for any  $\mathbf{w} \in \mathbb{R}^{n_i}$  we have

$$(3.7) \quad L_i^k(\mathbf{w}) - L_i^k(\mathbf{u}) \geq \frac{\delta}{2} (\|\mathbf{w} - \mathbf{u}\|^2 - \|\mathbf{w} - \mathbf{v}\|^2) + \frac{\rho}{2} \|\mathbf{A}_i(\mathbf{w} - \mathbf{u})\|^2 + \frac{\sigma\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2,$$

where  $L_i^k$  is given by

$$(3.8) \quad L_i^k(\mathbf{u}) = f_i(\mathbf{u}) + h_i(\mathbf{u}) + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2.$$

*Proof.* Adding  $h_i(\mathbf{u}) + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2$  to each side of the inequality (3.6) and rearranging, we obtain

$$\Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) - \frac{\sigma\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2 \geq L_i^k(\mathbf{u}).$$

Adding  $L_i^k(\mathbf{w})$  to each side of this inequality gives

$$(3.9) \quad L_i^k(\mathbf{w}) - L_i^k(\mathbf{u}) \geq L_i^k(\mathbf{w}) - \Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) + \frac{\sigma\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2.$$

Utilizing the convexity inequality  $f_i(\mathbf{w}) - f_i(\mathbf{v}) \geq \langle \nabla f_i(\mathbf{v}), \mathbf{w} - \mathbf{v} \rangle$ , we have

$$\begin{aligned} L_i^k(\mathbf{w}) - \Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) &\geq \frac{\rho}{2} (\|\mathbf{A}_i \mathbf{w} - \mathbf{b}_i^k - \boldsymbol{\lambda}^k / \rho\|^2 - \|\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k - \boldsymbol{\lambda}^k / \rho\|^2) \\ &\quad + \langle \nabla f_i(\mathbf{v}), \mathbf{w} - \mathbf{u} \rangle - \frac{\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2 + h_i(\mathbf{w}) - h_i(\mathbf{u}). \end{aligned}$$

Expand the smooth terms involving  $\mathbf{w}$  on the right side in a Taylor series around  $\mathbf{u}$  to obtain

$$L_i^k(\mathbf{w}) - \Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) \geq \langle \mathbf{g}_i^k, \mathbf{w} - \mathbf{u} \rangle + h_i(\mathbf{w}) - h_i(\mathbf{u}) + \frac{\rho}{2} \|\mathbf{A}_i(\mathbf{w} - \mathbf{u})\|^2 - \frac{\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2,$$

where  $\mathbf{g}_i^k = \nabla f_i(\mathbf{v}) + \rho \mathbf{A}_i^\top (\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho)$ . Since  $\Phi_i^k$  is the sum of smooth and nonsmooth terms, the first-order optimality condition for the minimizer  $\mathbf{u}$  of  $\Phi_i^k$  can be expressed

$$\langle \mathbf{g}_i^k + \delta(\mathbf{u} - \mathbf{v}), \mathbf{w} - \mathbf{u} \rangle + h_i(\mathbf{w}) \geq h_i(\mathbf{u})$$

for all  $\mathbf{w} \in \mathbb{R}^{n_i}$ , which implies that

$$\langle \mathbf{g}_i^k, \mathbf{w} - \mathbf{u} \rangle + h_i(\mathbf{w}) - h_i(\mathbf{u}) \geq -\delta \langle \mathbf{u} - \mathbf{v}, \mathbf{w} - \mathbf{u} \rangle$$

for all  $\mathbf{w} \in \mathbb{R}^{n_i}$ . Utilizing this inequality, we have

$$L_i^k(\mathbf{w}) - \Phi_i^k(\mathbf{u}, \mathbf{v}, \delta) \geq -\delta \langle \mathbf{u} - \mathbf{v}, \mathbf{w} - \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{A}_i(\mathbf{w} - \mathbf{u})\|^2 - \frac{\delta}{2} \|\mathbf{u} - \mathbf{v}\|^2.$$

Insert this in (3.9). Since

$$2\langle \mathbf{u} - \mathbf{v}, \mathbf{w} - \mathbf{u} \rangle + \|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{w} - \mathbf{v}\|^2 - \|\mathbf{w} - \mathbf{u}\|^2,$$

the proof is complete.  $\square$

We use Lemma 3.2 to establish a decay property that is key to the convergence analysis.

**LEMMA 3.3.** *Let  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$  be any solution/multiplier pair for (1.1)–(1.2), let  $\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k$ , and  $\boldsymbol{\lambda}^k$  be the iterates of the generalized BOSVS algorithm, and define*

$$E_k = \rho \|\mathbf{y}_+^k - \mathbf{x}_+^*\|_{\mathbf{P}}^2 + \frac{1}{\rho} \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 + \alpha \sum_{i=1}^m \delta_i^k \|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2,$$

where  $\mathbf{P} = \mathbf{M}\mathbf{H}^{-1}\mathbf{M}^\top$ . Then for  $k$  large enough that the monotonicity condition (3.4) holds for all  $i \in [1, m]$ , we have

$$E_k \geq E_{k+1} + c_1 \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 + c_2 \rho (\|\mathbf{y}_+^k - \mathbf{z}_+^k\|_{\mathbf{H}}^2 + \|\mathbf{A}\mathbf{z}^k - \mathbf{b}\|^2),$$

where  $c_1 = \sigma\alpha\delta_{\min}$  and  $c_2 = \alpha(1 - \alpha)$ .

*Proof.* By the inequality (3.7) of Lemma 3.2 with  $\mathbf{v} = \mathbf{x}_i^k$ ,  $\mathbf{w} = \mathbf{x}_i^*$ , and  $\mathbf{u} = \mathbf{z}_i^k$ , we have

$$\begin{aligned} L_i^k(\mathbf{x}_i^*) - L_i^k(\mathbf{z}_i^k) - \frac{\rho}{2} \|\mathbf{A}_i \mathbf{z}_{e,i}^k\|^2 &\geq \frac{\delta_i^k}{2} (\|\mathbf{z}_{e,i}^k\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma\delta_i^k}{2} \|\mathbf{z}_i^k - \mathbf{x}_i^k\|^2 \\ (3.10) \quad &= \frac{\delta_i^k}{2} (\|\mathbf{x}_{e,i}^{k+1}\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma\delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \end{aligned}$$

where  $\mathbf{x}_e^k = \mathbf{x}^k - \mathbf{x}^*$ ,  $\mathbf{z}_e^k = \mathbf{z}^k - \mathbf{x}^*$ , and  $\mathbf{z}^k = \mathbf{x}^{k+1}$  by Step 1b of generalized BOSVS. Since  $\mathbf{x}^*$  minimizes  $L(\cdot, \boldsymbol{\lambda}^*)$  and since the augmented Lagrangian is the sum of smooth and nonsmooth terms, the first-order optimality condition implies that for each  $i \in [1, m]$ ,

$$(3.11) \quad \langle \mathbf{g}_i^*, \mathbf{w} - \mathbf{x}_i^* \rangle + h_i(\mathbf{w}) - h_i(\mathbf{x}_i^*) \geq 0$$

for all  $\mathbf{w} \in \mathbb{R}^{n_i}$ , where  $\mathbf{g}_i^*$  is the gradient of the smooth part of the objective evaluated at  $\mathbf{x}_i^*$ :

$$\mathbf{g}_i^* = \nabla f_i(\mathbf{x}_i^*) + \rho \mathbf{A}_i^\top \left( \sum_{i=1}^m \mathbf{A}_i \mathbf{x}_i^* - \mathbf{b} + \boldsymbol{\lambda}^k / \rho \right).$$

We add  $L_i^k(\mathbf{x}_i^*) - L_i^k(\mathbf{z}_i^k) - \rho \|\mathbf{A}_i \mathbf{z}_{e,i}^k\|^2 / 2$  to both sides of (3.11) and take  $\mathbf{w} = \mathbf{z}_i^k$ . After much cancellation, we obtain the relation

$$\begin{aligned}
& L_i^k(\mathbf{x}_i^*) - L_i^k(\mathbf{z}_i^k) - \frac{\rho}{2} \|\mathbf{A}_i \mathbf{z}_{e,i}^k\|^2 \leq \\
& f_i(\mathbf{x}_i^*) - f_i(\mathbf{z}_i^k) + \nabla f_i(\mathbf{x}_i^*)^\top \mathbf{z}_{e,i}^k - \rho \left\langle \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho, \mathbf{A}_i \mathbf{z}_{e,i}^k \right\rangle \leq \\
(3.12) \quad & - \rho \left\langle \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho, \mathbf{A}_i \mathbf{z}_{e,i}^k \right\rangle,
\end{aligned}$$

where  $\mathbf{y}_e^k = \mathbf{y}^k - \mathbf{x}^*$ ,  $\boldsymbol{\lambda}_e^k = \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*$ , and the last inequality is due to the convexity of  $f_i$ . We combine this upper bound with the lower bound (3.10) to obtain

$$\begin{aligned}
& -\rho \left\langle \mathbf{A}_i \mathbf{z}_{e,i}^k, \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho \right\rangle \\
(3.13) \quad & \geq \frac{\delta_i^k}{2} (\|\mathbf{x}_{e,i}^{k+1}\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma \delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2.
\end{aligned}$$

Focusing on the left side of (3.13), observe that

$$\begin{aligned}
& \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k = \sum_{j=1}^m \mathbf{A}_j (\mathbf{z}_j^k - \mathbf{x}_j^*) + \sum_{j > i} \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) \\
(3.14) \quad & = \mathbf{A} \mathbf{z}^k - \mathbf{b} + \sum_{j > i} \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k)
\end{aligned}$$

since  $\mathbf{A} \mathbf{x}^* = \mathbf{b}$ . Let  $\tau_i^k$  denote the right side of (3.13):

$$\tau_i^k = \frac{\delta_i^k}{2} (\|\mathbf{x}_{e,i}^{k+1}\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma \delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2.$$

With this notation and with the simplification (3.14), (3.13) becomes

$$(3.15) \quad -\rho \left\langle \mathbf{A}_i \mathbf{z}_{e,i}^k, \mathbf{A} \mathbf{z}^k - \mathbf{b} + \boldsymbol{\lambda}_e^k / \rho + \sum_{j > i} \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) \right\rangle \geq \tau_i^k.$$

We will sum the inequality (3.15) over  $i$  between 1 and  $m$ . Since

$$\sum_{i=1}^m \mathbf{A}_i \mathbf{z}_{e,i}^k = \sum_{i=1}^m \mathbf{A}_i (\mathbf{z}_i^k - \mathbf{x}_i^*) = \mathbf{A} \mathbf{z}^k - \mathbf{b} := \mathbf{r}^k,$$

it follows that in (3.15),

$$(3.16) \quad \sum_{i=1}^m \langle \mathbf{A}_i \mathbf{z}_{e,i}^k, \mathbf{r}^k + \boldsymbol{\lambda}_e^k / \rho \rangle = \langle \mathbf{r}^k, \mathbf{r}^k + \boldsymbol{\lambda}_e^k / \rho \rangle.$$

Also, observe that

$$\sum_{j > i} \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) = \sum_{j=2}^m \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) - \sum_{j=2}^i \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k),$$



with the convention that the sum from  $j = 2$  to  $j = 1$  is 0. Take the inner product of this identity with  $\mathbf{A}_i \mathbf{z}_{e,i}^k$  and sum over  $i$  to obtain

$$(3.17) \quad \sum_{i=1}^m \left\langle \mathbf{A}_i \mathbf{z}_{e,i}^k, \sum_{j>i} \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) \right\rangle = \left\langle \mathbf{r}^k, \sum_{j=2}^m \mathbf{A}_j (\mathbf{y}_j^k - \mathbf{z}_j^k) \right\rangle - (\mathbf{z}_+^k - \mathbf{x}_+^*)^\top \mathbf{M} (\mathbf{y}_+^k - \mathbf{z}_+^k),$$

where  $\mathbf{M}$  is defined in (2.1). We sum (3.15) over  $i$  between 1 and  $m$  and utilize (3.16) and (3.17) to obtain

$$(3.18) \quad (\mathbf{y}_+^k - \mathbf{x}_+^*)^\top \mathbf{M} \mathbf{w} - \frac{1}{\rho} \left( \langle \mathbf{r}^k, \boldsymbol{\lambda}_e^k \rangle + \sum_{i=1}^m \tau_i^k \right) \geq \mathbf{w}^\top \mathbf{M} \mathbf{w} + \left\langle \mathbf{r}^k, \mathbf{r}^k + \sum_{j=2}^m \mathbf{A}_j \mathbf{w}_{j-1} \right\rangle,$$

where  $\mathbf{w} = \mathbf{y}_+^k - \mathbf{z}_+^k$ .

Observe that

$$\begin{aligned} \mathbf{w}^\top \mathbf{M} \mathbf{w} &= \frac{1}{2} \mathbf{w}^\top (\mathbf{M} + \mathbf{M}^\top) \mathbf{w} = \frac{1}{2} \mathbf{w}^\top (\mathbf{M} + \mathbf{M}^\top - \mathbf{H}) \mathbf{w} + \frac{1}{2} \mathbf{w}^\top \mathbf{H} \mathbf{w} \\ &= \frac{1}{2} \left\| \sum_{i=2}^m \mathbf{A}_i \mathbf{w}_{i-1} \right\|^2 + \frac{1}{2} \mathbf{w}^\top \mathbf{H} \mathbf{w} \end{aligned}$$

since  $(\mathbf{M} + \mathbf{M}^\top - \mathbf{H})_{ij} = \mathbf{A}_{i+1}^\top \mathbf{A}_{j+1}$  by the definition of  $\mathbf{M}$  and  $\mathbf{H}$ . With this substitution, the right side of (3.18) becomes a sum of squares:

$$\mathbf{w}^\top \mathbf{M} \mathbf{w} + \left\langle \mathbf{r}^k, \mathbf{r}^k + \sum_{j=2}^m \mathbf{A}_j \mathbf{w}_{j-1} \right\rangle = \frac{1}{2} \left( \mathbf{w}^\top \mathbf{H} \mathbf{w} + \|\mathbf{r}^k\|^2 + \left\| \mathbf{r}^k + \sum_{i=2}^m \mathbf{A}_i \mathbf{w}_{i-1} \right\|^2 \right).$$

Hence, it follows from (3.18) that

$$(3.19) \quad (\mathbf{y}_+^k - \mathbf{x}_+^*)^\top \mathbf{M} \mathbf{w} - \frac{1}{\rho} \left( \langle \mathbf{r}^k, \boldsymbol{\lambda}_e^k \rangle + \sum_{i=1}^m \tau_i^k \right) \geq \frac{1}{2} (\|\mathbf{w}\|_{\mathbf{H}}^2 + \|\mathbf{r}^k\|^2).$$

Let  $\mathbf{P} = \mathbf{M} \mathbf{H}^{-1} \mathbf{M}^\top$  and recall that  $\mathbf{w} = \mathbf{y}_+^k - \mathbf{z}_+^k$ . By the definition of  $\mathbf{y}^{k+1}$  and  $\boldsymbol{\lambda}^{k+1}$  in Step 3 of Algorithm 2.1, we have

$$\begin{aligned} &\|\mathbf{y}_+^k - \mathbf{x}_+^*\|_{\mathbf{P}}^2 - \|\mathbf{y}_+^{k+1} - \mathbf{x}_+^*\|_{\mathbf{P}}^2 + \frac{1}{\rho^2} (\|\boldsymbol{\lambda}_e^k\|^2 - \|\boldsymbol{\lambda}_e^{k+1}\|^2) = \\ &\|\mathbf{y}_+^k - \mathbf{x}_+^*\|_{\mathbf{P}}^2 - \|(\mathbf{y}_+^k - \mathbf{x}_+^*) - \alpha \mathbf{M}^{-\top} \mathbf{H} \mathbf{w}\|_{\mathbf{P}}^2 + \frac{1}{\rho^2} (\|\boldsymbol{\lambda}_e^k\|^2 - \|\boldsymbol{\lambda}_e^k + \alpha \rho \mathbf{r}^k\|^2) = \\ &2\alpha (\mathbf{y}_+^k - \mathbf{x}_+^*)^\top \mathbf{M} \mathbf{w} - \alpha^2 \|\mathbf{w}\|_{\mathbf{H}}^2 - \frac{2\alpha}{\rho} \langle \mathbf{r}^k, \boldsymbol{\lambda}_e^k \rangle - \alpha^2 \|\mathbf{r}^k\|^2. \end{aligned}$$

On the right side of this inequality, we utilize (3.19) multiplied by  $2\alpha$  to conclude that

$$(3.20) \quad \begin{aligned} &\|\mathbf{y}_+^k - \mathbf{x}_+^*\|_{\mathbf{P}}^2 - \|\mathbf{y}_+^{k+1} - \mathbf{x}_+^*\|_{\mathbf{P}}^2 + \frac{1}{\rho^2} (\|\boldsymbol{\lambda}_e^k\|^2 - \|\boldsymbol{\lambda}_e^{k+1}\|^2) - \frac{2\alpha}{\rho} \sum_{i=1}^m \tau_i^k \geq \\ &c_2 (\|\mathbf{y}_+^k - \mathbf{z}_+^k\|_{\mathbf{H}}^2 + \|\mathbf{r}^k\|^2) \end{aligned}$$

where  $c_2 = \alpha(1 - \alpha) > 0$  since  $\alpha \in (0, 1)$ . By the definition of  $\tau_i^k$  and the assumption that  $k$  is large enough that (3.4) holds for all  $i$ , it follows that

$$\begin{aligned} -\tau_i^k &= \frac{\delta_i^k}{2} (\|\mathbf{x}_{e,i}^k\|^2 - \|\mathbf{x}_{e,i}^{k+1}\|^2) - \frac{\sigma\delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \\ &\leq \frac{\delta_i^k}{2} \|\mathbf{x}_{e,i}^k\|^2 - \frac{\delta_i^{k+1}}{2} \|\mathbf{x}_{e,i}^{k+1}\|^2 - \frac{\sigma\delta_i^k}{2} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2. \end{aligned}$$

This bound for  $-\tau_i^k$  along with the inequality (3.20) and the definition of  $E^k$  complete the proof.  $\square$

The following theorem establishes the global convergence of generalized BOSVS.

**THEOREM 3.4.** *If  $\mathbf{x}^k$ ,  $\mathbf{y}^k$ , and  $\boldsymbol{\lambda}^k$  are iterates of the generalized BOSVS algorithm, then the  $\mathbf{x}^k$  and  $\mathbf{y}^k$  sequences converge to a common limit denoted  $\mathbf{x}^*$  and the  $\boldsymbol{\lambda}^k$  converge to a limit denoted  $\boldsymbol{\lambda}^*$  where  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$ .*

*Proof.* Let  $\bar{k}$  be chosen large enough that (3.4) holds for all  $k \geq \bar{k}$ . Since  $\mathbf{x}^{k+1} = \mathbf{z}^k$  in generalized BOSVS, it follows from Lemma 3.3 that for  $j \geq \bar{k}$  and  $p > 0$ , we have

$$(3.21) \quad E_j \geq E_{j+p} + c \sum_{k=j}^{j+p-1} (\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{y}_+^k - \mathbf{x}_+^{k+1}\|_{\mathbf{H}}^2 + \|\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\|^2),$$

where  $c = \min\{c_1, \rho c_2\} > 0$ . Let  $p$  tend to  $+\infty$  in (3.21). Since the columns of  $\mathbf{A}_i$  are linearly independent for  $i \geq 2$ ,  $\mathbf{H}$  is positive definite and

$$(3.22) \quad \lim_{k \rightarrow \infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = \lim_{k \rightarrow \infty} \|\mathbf{y}_+^k - \mathbf{x}_+^{k+1}\| = \lim_{k \rightarrow \infty} \|\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\| = 0.$$

By the definition of  $\mathbf{b}_i^k$ , we have

$$(3.23) \quad \mathbf{A}_i \mathbf{x}_i^{k+1} - \mathbf{b}_i^k = \sum_{j \leq i} \mathbf{A}_j \mathbf{x}_j^{k+1} + \sum_{j > i} \mathbf{A}_j \mathbf{y}_j^k - \mathbf{b}.$$

By (3.22),  $\mathbf{y}_+^k$  approach  $\mathbf{x}_+^{k+1}$ , and by (3.23) and (3.22),

$$(3.24) \quad \lim_{k \rightarrow \infty} (\mathbf{A}_i \mathbf{x}_i^{k+1} - \mathbf{b}_i^k) = \lim_{k \rightarrow \infty} \mathbf{A}\mathbf{x}^{k+1} - \mathbf{b} = \mathbf{0}$$

for all  $i \in [1, m]$ .

By the definition of  $E_k$  in Lemma 3.3, we see that the iterates  $\boldsymbol{\lambda}^k$  and  $\mathbf{x}^k$  are uniformly bounded. Hence, there exist limits  $\boldsymbol{\lambda}^*$  and  $\mathbf{x}^*$ , and an infinite sequence  $\mathcal{K} \subset \{1, 2, \dots\}$  such that  $\boldsymbol{\lambda}^k$  and  $\mathbf{x}^k$  for  $k \in \mathcal{K}$  converge to  $\boldsymbol{\lambda}^*$  and  $\mathbf{x}^*$  respectively. By the first relation in (3.22),  $\mathbf{x}^{k+1}$  also converges to  $\mathbf{x}^*$  for  $k \in \mathcal{K}$ . In Step 1b of generalized BOSVS, we have

$$\mathbf{x}_i^{k+1} = \arg \min \{ \Phi_i^k(\mathbf{u}, \mathbf{x}_i^k, \delta_i^k) : \mathbf{u} \in \mathbb{R}^{n_i} \}.$$

The first-order optimality conditions for  $\mathbf{x}_i^{k+1}$  are

$$(3.25) \quad \langle \mathbf{g}_i^k, \mathbf{u} - \mathbf{x}_i^{k+1} \rangle + h_i(\mathbf{u}) \geq h_i(\mathbf{x}_i^{k+1})$$

for all  $\mathbf{u} \in \mathbb{R}^{n_i}$ , where  $\mathbf{g}_i^k$  is the gradient of the smooth part of the objective evaluated at  $\mathbf{x}_i^{k+1}$ :

$$\mathbf{g}_i^k = \nabla f_i(\mathbf{x}_i^k) + \rho \mathbf{A}_i^\top (\mathbf{A}_i \mathbf{x}_i^{k+1} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho) + \delta_i^k (\mathbf{x}_i^{k+1} - \mathbf{x}_i^k).$$

As  $k \in \mathcal{K}$  tends to infinity,  $\nabla f_i(\mathbf{x}_i^k)$  approaches  $\nabla f_i(\mathbf{x}^*)$  since  $\nabla f_i$  is Lipschitz continuous,  $\mathbf{A}_i \mathbf{x}_i^{k+1} - \mathbf{b}_i^k$  approaches  $\mathbf{0}$  by (3.24), and  $\delta_i^k(\mathbf{x}_i^{k+1} - \mathbf{x}_i^k)$  approaches  $\mathbf{0}$  by (3.22) and the uniform bounded (3.4) for  $\delta_i^k$ . Consequently, we have

$$(3.26) \quad \lim_{k \in \mathcal{K}} \mathbf{g}_i^k = \nabla f_i(\mathbf{x}^*) + \mathbf{A}_i^\top \boldsymbol{\lambda}^*.$$

Let  $k \in \mathcal{K}$  tend to  $+\infty$  in (3.25). By (3.26) and the lower semicontinuity of  $h_i$ , we deduce that

$$\langle \nabla f_i(\mathbf{x}_i^*) + \mathbf{A}_i^\top \boldsymbol{\lambda}^*, \mathbf{u} - \mathbf{x}_i^* \rangle + h_i(\mathbf{u}) \geq h_i(\mathbf{x}_i^*)$$

for all  $\mathbf{u} \in \mathbb{R}^{n_i}$ . Therefore,  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}^*$  satisfy the first-order optimality condition (1.3). By the last relation in (3.22), it follows that  $\mathbf{A}\mathbf{x}^* = \mathbf{b}$  and  $\mathbf{x}^*$  is feasible in (1.1). By the convexity of  $f_i$  and  $h_i$ ,  $\mathbf{x}^*$  is a solution of (1.1)–(1.2) and  $\boldsymbol{\lambda}^*$  is an associated multiplier for the linear constraint.

Since  $\mathbf{x}^{k+1}$  converges to  $\mathbf{x}^*$  for  $k \in \mathcal{K}$ , the second relation in (3.22) implies that  $\mathbf{y}_+^k$  converges to  $\mathbf{x}_+^*$  for  $k \in \mathcal{K}$ . In Lemma 3.3, we use the specific limits  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}^*$  associated with  $k \in \mathcal{K}$ . Hence,  $E_k$  tends to 0 for  $k \in \mathcal{K}$ . It follows from (3.21) that the entire  $E_k$  sequence tends to 0. By the definition of  $E_k$ , we deduce that the entire  $(\mathbf{x}^k, \mathbf{y}_+^k, \boldsymbol{\lambda}^k)$  sequence converges  $(\mathbf{x}^*, \mathbf{x}_+^*, \boldsymbol{\lambda}^*)$ . Since  $\mathbf{y}_1^k = \mathbf{x}_1^k$  for each  $k$ , where  $\mathbf{x}_1^k$  converges to  $\mathbf{x}_1^*$ , we conclude that  $\mathbf{y}^k$  converges to  $\mathbf{x}^*$ . This completes the proof.  $\square$

**4. Multistep BOSVS.** For the template given by Algorithm 2.1, we only need to assume that the columns of  $\mathbf{A}_i$  are linearly independent for  $i \geq 2$  since only these columns enter into the matrix  $\mathbf{M}$  which is inverted in Step 3. For generalized BOSVS, this assumption was sufficient of convergence. On the other hand, for both multistep and accelerated BOSVS, strong convexity of the augmented Lagrangian with respect to each of the variables  $\mathbf{x}_i$  is needed in the analysis. Since it has already been assumed that the columns of  $\mathbf{A}_i$  are linearly independent for  $i \geq 2$ , we will simply strengthen this assumption to require, henceforth, that the columns of  $\mathbf{A}_i$  are linearly independent for every  $i$ . This ensures strong convexity of the augmented Lagrangian  $L$  with respect to each of the variables  $\mathbf{x}_i$ .

The inner loop for the multistep BOSVS algorithm appears in Algorithm 4.1. In

---

**Initialize:**  $\mathbf{u}_i^0 = \mathbf{x}_i^k$

**For**  $l = 1, 2, \dots$

1a. Choose  $\delta_0^l \in [\delta_{\min}, \delta_{\max}]$ .

1b. Set  $\delta^l = \eta^j \delta_0^l$ , where  $j \geq 0$  is the smallest integer such that

$$f_i(\mathbf{u}_i^{l-1}) + \langle \nabla f_i(\mathbf{u}_i^{l-1}), \mathbf{u}_i^l - \mathbf{u}_i^{l-1} \rangle + \frac{(1-\sigma)\delta^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \geq f_i(\mathbf{u}_i^l),$$

where  $\mathbf{u}_i^l = \arg \min \{ \Phi_i^k(\mathbf{u}, \mathbf{u}_i^{l-1}, \delta_i^k) : \mathbf{u} \in \mathbb{R}^{n_i} \}$ .

1c. If  $\gamma^l := \sum_{j=1}^l 1/\delta^j \geq \Gamma_i^{k-1}$  and  $\|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|/\sqrt{\gamma^l} \leq \psi(e^{k-1})$ , break.

**Next**

1d. Set  $\mathbf{z}_i^k = \left( \sum_{j=1}^l \mathbf{u}_i^j / \delta^j \right) / \gamma^l$ ,  $r_i^k = (1/\gamma^l) \sum_{j=1}^l \|\mathbf{u}_i^j - \mathbf{u}_i^{j-1}\|^2$ ,  
 $\mathbf{x}_i^{k+1} = \mathbf{u}_i^l$ , and  $\Gamma_i^k = \gamma^l$ .

---

ALG. 4.1. Inner loop in Step 1 of Algorithm 2.1 for the multistep BOSVS scheme.

generalized BOSVS, the iteration is given by  $\mathbf{x}_i^{k+1} = \arg \min \{ \Phi_i^k(\mathbf{u}, \mathbf{x}_i^k, \delta_i^k) : \mathbf{u} \in \mathbb{R}^{n_i} \}$

where  $\delta_i^k$  is determined by a line search process. In the multistep BOSVS algorithm, this single minimization is replaced by the recurrence

$$\mathbf{u}_i^l = \arg \min \{ \Phi_i^k(\mathbf{u}, \mathbf{u}_i^{l-1}, \delta_i^k) : \mathbf{u} \in \mathbb{R}^{n_i} \},$$

where  $\mathbf{u}_i^0 = \mathbf{x}_i^k$ . By converting the single minimization into a recurrence, we hope to achieve a better minimizer of the augmented Lagrangian. In generalized BOSVS, the convergence relies on a careful choice of  $\delta_i^k$  using both the BB-formula and a safeguarding technique. In multistep BOSVS, these restrictions on  $\delta_i^k$  are replaced in Step 1c by a condition related to the accuracy of the iterates. In this step,  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  denotes any function satisfying  $\psi(0) = 0$  and  $\psi(s) > 0$  for  $s > 0$  with  $\psi$  continuous at  $s = 0$ . For example,  $\psi(t) = t$ .

REMARK 4.1. *Computationally, it is not necessary to store the  $\mathbf{u}_i^l$  sequence to evaluate  $\mathbf{z}_i^k$  in Step 1d. For example, in Step 1c we could introduce a sequence*

$$\mathbf{a}_i^l = (1 - \alpha^l) \mathbf{a}_i^{l-1} + \alpha^l \mathbf{u}_i^l, \quad \text{where } \alpha^l = 1/(\delta^l \gamma^l) \text{ and } \mathbf{a}_i^0 = \mathbf{x}_i^0,$$

and in Step 1d, we would set  $\mathbf{z}_i^k = \mathbf{a}_i^l$ . Note that  $0 < \alpha^l \leq 1$  due to the form of  $\gamma^l$  in Step 1c.

Since  $\eta > 1$ , the line search in Step 1b of multistep BOSVS terminates in a finite number of iterations and the final  $\delta^l$  has exactly the same bounds (3.3) as that of generalized BOSVS. Since  $\delta^l$  is uniformly bounded, it follows that the condition  $\Gamma_i^k \geq \Gamma_i^{k-1}$  of Step 1c is fulfilled for  $l$  sufficiently large. In the numerical experiments for multistep BOSVS in Section 6,  $\delta_0^l$  is given by the safeguarded BB choice of generalized BOSVS.

Let us first observe that when  $e^k = 0$ , we have reached a solution of (1.1)–(1.2).

LEMMA 4.1. *If  $e^k = 0$  in the multistep BOSVS algorithm, then  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{y}^k = \mathbf{z}^k$  solves (1.1)–(1.2) and  $(\mathbf{x}^k, \boldsymbol{\lambda}^k) \in \mathcal{W}^*$ .*

*Proof.* If  $e^k = 0$ , then  $r_i^k = 0$  for each  $i$ . It follows that  $\mathbf{x}_i^k = \mathbf{u}_i^0 = \mathbf{u}_i^1 = \dots = \mathbf{u}_i^l$ . By Step 1d,  $\mathbf{u}_i^l = \mathbf{x}_i^{k+1} = \mathbf{z}_i^k$ . Consequently, we have  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{z}^k$ . Since all three algorithms in this paper share Algorithm 2.1, the remainder of the proof is exactly as in Lemma 3.1.  $\square$

The following inequality is based on Lemma 3.2.

LEMMA 4.2. *In multistep BOSVS, we have*

$$(4.1) \quad \nu_i \rho \|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\|^2 + \frac{\sigma}{\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \leq \frac{\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2}{\Gamma_i^k},$$

for each  $i \in [1, m]$ , where  $l_i^k$  is the terminating value of  $l$  at iteration  $k$ ,  $\nu_i > 0$  is the smallest eigenvalue of  $\mathbf{A}_i^T \mathbf{A}_i$ , and

$$(4.2) \quad \bar{\mathbf{x}}_i^k = \arg \min \{ L_i^k(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^{n_i} \}$$

with  $L_i^k$  defined in (3.8).

*Proof.* By Lemma 3.2, we have

$$(4.3) \quad L_i^k(\mathbf{w}) - L_i^k(\mathbf{u}_i^l) \geq \frac{\delta^l}{2} (\|\mathbf{w} - \mathbf{u}_i^l\|^2 - \|\mathbf{w} - \mathbf{u}_i^{l-1}\|^2) + \frac{\rho}{2} \|\mathbf{A}_i(\mathbf{w} - \mathbf{u}_i^l)\|^2 \\ + \frac{\sigma \delta^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2$$

for any  $\mathbf{w} \in \mathbb{R}^{n_i}$ . We take  $\mathbf{w} = \bar{\mathbf{x}}_i^k$ . Since  $L_i^k(\mathbf{u}_i^l) - L_i^k(\bar{\mathbf{x}}_i^k) \geq 0$ , we have

$$(4.4) \quad \frac{\rho}{\delta^l} \|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{u}_i^l)\|^2 + \sigma \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \leq \|\bar{\mathbf{x}}_i^k - \mathbf{u}_i^{l-1}\|^2 - \|\bar{\mathbf{x}}_i^k - \mathbf{u}_i^l\|^2.$$

Summing this inequality for  $l$  between 1 and  $l_i^k$  gives

$$(4.5) \quad \rho \sum_{l=1}^{l_i^k} \frac{1}{\delta^l} \|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{u}_i^l)\|^2 + \sigma \sum_{l=1}^{l_i^k} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \leq \|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^k\|^2.$$

Since the quadratic  $\|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{u})\|^2$  is a convex function of  $\mathbf{u}$ , it follows from Jensen's inequality that

$$\sum_{l=1}^{l_i^k} \frac{1}{\delta^l} \|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{u}_i^l)\|^2 \geq \Gamma_i^k \|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{z}_i^k)\|^2 \geq \Gamma_i^k \nu_i \|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\|^2,$$

where  $\nu_i > 0$  is the smallest eigenvalue of  $\mathbf{A}_i^\top \mathbf{A}_i$ . Combine this with (4.5) to obtain (4.1).  $\square$

REMARK 4.2.  $L_i^k$  is strongly convex since it is the sum of convex functions and a strongly convex quadratic  $\langle \mathbf{A}_i \mathbf{u}, \mathbf{A}_i \mathbf{u} \rangle$ ; consequently, the minimizer  $\bar{\mathbf{x}}_i^k$  exists. Due to the upper bound (3.3) for  $\delta^l$  in multistep BOSVS,  $\gamma^l$  grows linearly in  $l$ . Hence, for the inner loop of multistep BOSVS, (4.1) implies that  $\|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\| = O(1/\sqrt{l_i^k})$ . By (4.3), the objective values satisfy  $L_i^k(\mathbf{z}_i^k) - L_i^k(\bar{\mathbf{x}}_i^k) = O(1/l_i^k)$ ; to see this, divide (4.3) by  $\delta^l$ , sum over  $l$  between 1 and  $l_i^k$ , and apply Jensen's inequality twice, to the terms involving  $L(\mathbf{u}^l)$  and to the terms involving  $\mathbf{A}_i$ .

As a consequence of Lemma 4.2, we show that the stopping conditions in Step 1c of multistep BOSVS are satisfied for a finite  $l$ .

COROLLARY 4.3. If  $e^{k-1} > 0$  in Step 1d of multistep BOSVS, then the stopping condition in Step 1c is fulfilled when  $l$  is sufficiently large.

*Proof.* Since  $\delta^l$  in multistep BOSVS has the same upper bound (3.3) as generalized BOSVS, it follows that  $\Gamma_i^k$  in step 1c of multistep BOSVS tends to infinity as  $l$  tends to infinity. By Lemma 4.2, the iteration difference  $\|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|$  tends to zero as  $l_i^k$  grows. Hence, both conditions in Step 1c are satisfied for  $l$  sufficiently large when  $\psi(e^{k-1}) > 0$ .  $\square$

Similar to generalized BOSVS, the key to the convergence of multistep BOSVS is a decay property for the iterates. The analogue of Lemma 3.3 for multistep BOSVS is the following result.

LEMMA 4.4. Let  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$  be any solution/multiplier pair for (1.1)–(1.2), let  $\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}_i^k$ , and  $\boldsymbol{\lambda}^k$  be the iterates of the multistep BOSVS algorithm, let  $l_i^k$  be the terminating value of  $l$  at iteration  $k$ , and define

$$E_k = \rho \|\mathbf{y}_+^k - \mathbf{x}_+^*\|_{\mathbf{P}}^2 + \frac{1}{\rho} \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 + \alpha \sum_{i=1}^m \frac{\|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2}{\Gamma_i^k},$$

where  $\mathbf{P} = \mathbf{M}\mathbf{H}^{-1}\mathbf{M}^\top$ . Then for all  $k$  and for all  $i \in [1, m]$ , we have

$$E_k \geq E_{k+1} + c_1 \sum_{i=1}^m \sum_{l=1}^{l_i^k} \frac{\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2}{\Gamma_i^k} + c_2 \rho (\|\mathbf{y}_+^k - \mathbf{z}_+^k\|_{\mathbf{H}}^2 + \|\mathbf{A}\mathbf{z}^k - \mathbf{b}\|^2),$$

where  $c_1 = \sigma\alpha$  and  $c_2 = \alpha(1 - \alpha)$ .

*Proof.* We put  $\mathbf{w} = \mathbf{x}_i^*$  in (4.3) to obtain

$$\frac{L_i^k(\mathbf{x}_i^*) - F_i^k(\mathbf{u}_{i,k}^l)}{\delta^l} \geq \frac{1}{2}(\|\mathbf{x}_i^* - \mathbf{u}_{i,k}^l\|^2 - \|\mathbf{x}_i^* - \mathbf{u}_{i,k}^{l-1}\|^2) + \frac{\sigma}{2}\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2,$$

where  $F_i^k(\mathbf{u}_{i,k}^l) = L_i^k(\mathbf{u}_{i,k}^l) + (\rho/2)\|\mathbf{A}_i(\mathbf{u}_{i,k}^l - \mathbf{x}_i^*)\|^2$ . Summing this inequality over  $l$  yields

$$(4.6) \quad \sum_{l=1}^{l_i^k} \left( \frac{L_i^k(\mathbf{x}_i^*) - F_i^k(\mathbf{u}_{i,k}^l)}{\delta^l} \right) \geq \frac{1}{2}(\|\mathbf{x}_i^* - \mathbf{u}_{i,k}^{l_i^k}\|^2 - \|\mathbf{x}_i^* - \mathbf{u}_{i,k}^0\|^2) + \frac{\sigma}{2} \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2.$$

Since  $F_i^k$  is convex, it follows from Jensen's inequality and the definition of  $\Gamma_i^k$  and  $\mathbf{z}_i^k$  in Step 1c of multistep BOSVS that

$$(4.7) \quad \frac{1}{\Gamma_i^k} \sum_{l=1}^{l_i^k} \frac{1}{\delta^l} F_i^k(\mathbf{u}_{i,k}^l) \geq F_i^k \left( \frac{1}{\Gamma_i^k} \sum_{l=1}^{l_i^k} \frac{1}{\delta^l} \mathbf{u}_{i,k}^l \right) = F_i^k(\mathbf{z}_i^k).$$

Substitute  $\mathbf{x}_i^{k+1} = \mathbf{u}_{i,k}^{l_i^k}$  and  $\mathbf{x}_i^k = \mathbf{u}_{i,k}^0$  in (4.6) and use (4.7) to obtain

$$(4.8) \quad L_i^k(\mathbf{x}_i^*) - F_i^k(\mathbf{z}_i^k) \geq \frac{1}{2\Gamma_i^k}(\|\mathbf{x}_i^{k+1}\|^2 - \|\mathbf{x}_i^k\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2,$$

where  $\mathbf{x}_{e,i}^k = \mathbf{x}_i^k - \mathbf{x}_i^*$ . By (3.12), we have the upper bound

$$L_i^k(\mathbf{x}_i^*) - F_i^k(\mathbf{z}_i^k) \leq -\rho \left\langle \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho, \mathbf{A}_i \mathbf{z}_{e,i}^k \right\rangle.$$

Combining lower and upper bounds gives

$$(4.9) \quad -\rho \left\langle \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho, \mathbf{A}_i \mathbf{z}_{e,i}^k \right\rangle \geq \frac{1}{2\Gamma_i^k}(\|\mathbf{x}_{e,i}^{k+1}\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2,$$

which is the same as (3.13) but with the following exchanges:

$$\delta_i^k \longleftrightarrow 1/\Gamma_i^k \quad \text{and} \quad \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \longleftrightarrow \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2.$$

Except for these adjustments, the remainder of the proof is the same as the proof of Lemma 3.3, starting with equation (3.14).  $\square$

Using Lemma 4.4, we can now prove the convergence of multistep BOSVS. The analysis parallels that of Theorem 3.4. To facilitate the analysis, we recall the definition and some properties of the prox function. For any closed convex real-valued function  $h$ ,

$$\text{prox}_h(\mathbf{v}) = \arg \min \left\{ h(\mathbf{u}) + \frac{1}{2} \|\mathbf{v} - \mathbf{u}\|^2 : \mathbf{u} \in \text{dom}(h) \right\}.$$

As shown in [32, p. 340], the prox function is nonexpansive:

$$\|\text{prox}_h(\mathbf{v}_1) - \text{prox}_h(\mathbf{v}_2)\| \leq \|\mathbf{v}_1 - \mathbf{v}_2\|.$$

Moreover, if  $g$  is a differentiable convex function and

$$(4.10) \quad \mathbf{u}^* = \arg \min_{\mathbf{u}} g(\mathbf{u}) + h(\mathbf{u}),$$

then it follows from the first-order optimality conditions for  $\mathbf{u}^*$  that

$$(4.11) \quad \mathbf{u}^* = \text{prox}_h(\mathbf{u}^* - \nabla g(\mathbf{u}^*)).$$

Conversely, if (4.11) holds, then so does (4.10). Hence, these relations are equivalent. These properties will be used in the convergence analysis of multistep BOSVS.

**THEOREM 4.5.** *If multistep BOSVS performs an infinite number of iterations generating iterates  $\mathbf{y}^k$ ,  $\mathbf{z}^k$ , and  $\boldsymbol{\lambda}^k$ , then the sequences  $\mathbf{y}^k$  and  $\mathbf{z}^k$  both approach a common limit  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}^k$  approaches a limit  $\boldsymbol{\lambda}^*$  where  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$ .*

*Proof.* For any  $p > 0$ , we sum the decay property of Lemma 4.4 to obtain

$$(4.12) \quad E_j \geq E_{j+p} + c \sum_{k=j}^{j+p-1} \left( \|\mathbf{y}_+^k - \mathbf{z}_+^k\|_{\mathbf{H}}^2 + \|\mathbf{A}\mathbf{z}^k - \mathbf{b}\|^2 + \sum_{i=1}^m \sum_{l=1}^{l_i^k} \frac{\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2}{\Gamma_i^k} \right),$$

where  $c = \min\{c_1, \rho c_2\} > 0$ . Let  $p$  tend to  $+\infty$ . Since  $\mathbf{H}$  is positive definite, and the  $\Gamma_i^k$  are monotone nondecreasing as a function of  $k$ , it follows from (4.12) that

$$(4.13) \quad \lim_{k \rightarrow \infty} \|\mathbf{y}_+^k - \mathbf{z}_+^k\| = 0 = \lim_{k \rightarrow \infty} \|\mathbf{A}\mathbf{z}^k - \mathbf{b}\|.$$

Moreover, by the definition of  $E_k$  in Lemma 4.4,  $\mathbf{y}_+^k$  and  $\boldsymbol{\lambda}^k$  are bounded sequences, and by the first equation in (4.13),  $\mathbf{z}_+^k$  is also a bounded sequence. The second equation in (4.13) is equivalent to

$$\lim_{k \rightarrow \infty} \left\| \mathbf{A}_1 \mathbf{z}_1^k - \left( \mathbf{b} - \sum_{i=2}^m \mathbf{A}_i \mathbf{z}_i^k \right) \right\| = 0.$$

Since  $\mathbf{z}_+^k$  is bounded and the columns of  $\mathbf{A}_1$  are linearly independent,  $\mathbf{z}_1$  is bounded. Hence, both  $\mathbf{z}^k$  and  $\boldsymbol{\lambda}^k$  are bounded sequences, and there exist an infinite sequence  $\mathcal{K} \subset \{1, 2, \dots\}$  and limits  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}^*$  such that

$$(4.14) \quad \lim_{k \in \mathcal{K}} \mathbf{z}^k = \mathbf{x}^* \quad \text{and} \quad \lim_{k \in \mathcal{K}} \boldsymbol{\lambda}^k = \boldsymbol{\lambda}^*.$$

By the first equation in (4.13), we have

$$(4.15) \quad \lim_{k \in \mathcal{K}} \mathbf{y}_+^k = \mathbf{x}_+^*.$$

By the second equation in (4.13),  $\mathbf{Ax}^* = \mathbf{b}$ . Consequently, by (4.14) and (4.15),

$$(4.16) \quad \lim_{k \in \mathcal{K}} (\mathbf{A}_i \mathbf{z}_i^k - \mathbf{b}_i^k) = \lim_{k \in \mathcal{K}} \left( \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_j^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_j^k - \mathbf{b} \right) = \mathbf{Ax}^* - \mathbf{b} = \mathbf{0}$$

for all  $i \in [1, m]$ .

The decay property (4.12) also implies that for each  $i$ ,

$$(4.17) \quad \lim_{k \rightarrow \infty} r_i^k = \lim_{k \rightarrow \infty} \frac{1}{\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2 = 0.$$

Combine this with (4.13) to conclude that

$$(4.18) \quad \lim_{k \rightarrow \infty} e^k = \lim_{k \rightarrow \infty} \psi(e^k) = 0.$$

The remainder of the proof is partitioned into two cases depending on whether the monotone nondecreasing sequence  $\Gamma_i^k$  either approaches a finite limit, or tends to infinity.

**Case 1.** For some  $i$ ,  $\Gamma_i^k$  approaches a finite limit. Due to the upper bound (3.3) for  $\delta^l$  in Step 1b of multistep BOSVS, we conclude that  $l_i^k$  is uniformly bounded. By (4.17),  $\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|$  approaches zero, where the convergence is uniform in  $k$  and  $l \in [1, l_i^k]$ . Since  $\mathbf{u}_{i,k}^0 = \mathbf{x}_i^k$ , the triangle inequality and the uniform upper bound for  $l_i^k$  imply that  $\|\mathbf{x}_i^k - \mathbf{u}_{i,k}^l\|$  approaches zero, where the convergence is uniform in  $k$  and  $l \in [1, l_i^k]$ . Since  $\mathbf{z}_i^k$  is a convex combination of  $\mathbf{u}_{i,k}^l$  for  $0 \leq l \leq l_i^k$  with  $l_i^k$  uniformly bounded and  $\|\mathbf{x}_i^k - \mathbf{u}_{i,k}^l\|$  approaching zero, it follows that  $\|\mathbf{z}_i^k - \mathbf{x}_i^k\|$  approaches zero. We summarize these observations in the relation

$$(4.19) \quad \lim_{k \rightarrow \infty} \|\mathbf{z}_i^k - \mathbf{x}_i^k\| = \lim_{k \rightarrow \infty} \|\mathbf{z}_i^k - \mathbf{u}_{i,k}^0\| = \lim_{k \rightarrow \infty} \|\mathbf{z}_i^k - \mathbf{u}_{i,k}^1\| = 0.$$

In multistep BOSVS,  $\mathbf{u}_{i,k}^1$  minimizes  $\Phi_i(\cdot, \mathbf{u}_i^0, \delta_i^k)$ . Identify  $g$  in (4.10) with the smooth terms in  $\Phi_i$ . By (4.11), we have

$$\mathbf{u}_{i,k}^1 = \text{prox}_{h_i}(\mathbf{u}_{i,k}^1 - \nabla f_i(\mathbf{u}_{i,k}^0) - \delta_i^k(\mathbf{u}_{i,k}^1 - \mathbf{u}_{i,k}^0) - \rho \mathbf{A}_i^\top (\mathbf{A}_i \mathbf{u}_{i,k}^1 - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho)).$$

Let us now take the limit as  $k$  tends to infinity with  $k \in \mathcal{K}$ . By (4.14),  $\mathbf{z}_i^k$  approaches  $\mathbf{x}_i^*$ . By (4.19) both  $\mathbf{u}_{i,k}^0$  and  $\mathbf{u}_{i,k}^1$  approach  $\mathbf{z}_i^k$ , and by (4.16)  $\mathbf{A}_i \mathbf{u}_{i,k}^1 - \mathbf{b}_i^k$  approaches zero. Since the prox function and  $\nabla f_i$  are both Lipschitz continuous, we deduce that in the limit, as  $k$  tends to infinity with  $k \in \mathcal{K}$ ,

$$\mathbf{x}_i^* = \text{prox}_{h_i}(\mathbf{x}_i^* - \nabla f_i(\mathbf{x}_i^*) - \mathbf{A}_i^\top \boldsymbol{\lambda}^*).$$

Again, by (4.10), we have

$$(4.20) \quad \mathbf{x}_i^* = \arg \min \{f_i(\mathbf{u}) + h_i(\mathbf{u}) + \langle \boldsymbol{\lambda}^*, \mathbf{A}_i \mathbf{u} \rangle : \mathbf{u} \in \mathbb{R}^{n_i}\}.$$

If this were to hold for all  $i \in [1, m]$ , then it would follow that

$$(4.21) \quad \mathbf{x}^* = \arg \min \{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*) : \mathbf{x} \in \mathbb{R}^n\}.$$

Since  $\mathbf{Ax}^* = \mathbf{b}$ , we conclude that  $\mathbf{x}^*$  is an optimal solution of (1.1)–(1.2), and  $\boldsymbol{\lambda}^*$  is an associated multiplier. To show that (4.20) holds for all  $i$ , we need to consider the situation where  $\Gamma_i^k$  tends to infinity.



**Case 2.** Suppose that  $\Gamma_i^k$  approaches infinity. Let  $\bar{\mathbf{x}}_i^k$  be the minimizer of  $L_i^k$  defined in (3.8). Observe that minimizing  $L_i^k(\mathbf{u})$  over  $\mathbf{u} \in \mathbb{R}^{n_i}$  is equivalent to minimizing a sum of the form  $g(\mathbf{u}) + h(\mathbf{u}) + \langle \mathbf{u}, \mathbf{c}^k \rangle$  where  $h$  corresponds to  $h_i$ ,  $\mathbf{c}^k = \mathbf{A}_i^\top (\boldsymbol{\lambda}^k - \rho \mathbf{b}_i^k)$ , and  $g(\mathbf{u}) = f_i(\mathbf{u}) + 0.5\rho \|\mathbf{A}_i \mathbf{u}\|^2$ . Note that  $g$  is smooth and satisfies a strong convexity condition

$$(4.22) \quad (\mathbf{u} - \mathbf{v})^\top (\nabla g(\mathbf{u}) - \nabla g(\mathbf{v})) \geq \rho \nu_i \|\mathbf{u} - \mathbf{v}\|^2,$$

where  $\nu_i > 0$  is the smallest eigenvalue of  $\mathbf{A}_i^\top \mathbf{A}_i$ . By the strong convexity of  $L_i^k$ , it has a unique minimizer, and from the first-order optimality conditions and the strong convexity condition (4.22), we obtain the bound

$$(4.23) \quad \|\bar{\mathbf{x}}_i^j - \bar{\mathbf{x}}_i^k\| \leq \|\mathbf{c}^j - \mathbf{c}^k\| / (\rho \nu_i).$$

Since  $\mathbf{z}^k$ ,  $\mathbf{y}_+^k$ , and  $\boldsymbol{\lambda}^k$  are bounded sequences, it follows that  $\bar{\mathbf{x}}_i^k$  is a bounded sequence. For  $k \in \mathcal{K}$ , the sequences  $\mathbf{z}^k$ ,  $\mathbf{y}_+^k$ , and  $\boldsymbol{\lambda}^k$  converge to  $\mathbf{x}^*$ ,  $\mathbf{x}_+^*$ , and  $\boldsymbol{\lambda}^*$  respectively, which implies that

$$(4.24) \quad \mathbf{c}^* = \lim_{k \in \mathcal{K}} \mathbf{c}^k = \mathbf{A}_i^\top \left[ \boldsymbol{\lambda}^* - \rho \left( \mathbf{b} - \sum_{j \neq i} \mathbf{A}_j \mathbf{x}_j^* \right) \right] = \mathbf{A}_i^\top [\boldsymbol{\lambda}^* - \rho \mathbf{A}_i \mathbf{x}_i^*],$$

where the last equality is due to the identity  $\mathbf{A} \mathbf{x}^* = \mathbf{b}$ . Consequently, by (4.23),  $\bar{\mathbf{x}}_i^k$  for  $k \in \mathcal{K}$  forms a Cauchy sequence which approaches a limit. We use the stopping condition to determine the limit.

Let us insert  $l = l_i^k$  and  $\mathbf{u}_i^l = \mathbf{x}_i^{k+1}$  in the inequality (4.4). By the linear independence of the columns of  $\mathbf{A}_i$  and the upper bound (3.3) for  $\delta^l$ , there exists  $\beta > 0$  such that

$$\begin{aligned} \beta \|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}\|^2 &\leq \frac{\rho}{\delta^l} \|\mathbf{A}_i(\bar{\mathbf{x}}_i^k - \mathbf{u}_{i,k}^l)\|^2 \leq \|\bar{\mathbf{x}}_i^k - \mathbf{u}_{i,k}^{l-1}\|^2 - \|\bar{\mathbf{x}}_i^k - \mathbf{u}_{i,k}^l\|^2 \\ &= 2\langle \bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}, \mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1} \rangle + \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \\ &\leq 2\|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}\| \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\| + \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2. \end{aligned}$$

We move the  $\|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}\|$  term on the right side of this inequality to the left side and complete the square to obtain the relation

$$\|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}\| \leq \frac{\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|}{\beta} \left( 1 + \sqrt{\beta + 1} \right).$$

Square this inequality and divide by  $\Gamma_i^k$  to get

$$\frac{\|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^{k+1}\|^2}{\Gamma_i^k} \leq \frac{\|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2}{\Gamma_i^k} \left( \frac{1 + \sqrt{\beta + 1}}{\beta^2} \right)^2.$$

Since  $l = l_i^k$ , it follows from the stopping condition of Step 1c and from (4.18) that the right of this inequality approaches zero as  $k$  tends to infinity. Earlier we showed that  $\bar{\mathbf{x}}_i^k$  is a bounded sequence. Since  $\Gamma_i^k$  tends to infinity in Case 2, and  $\bar{\mathbf{x}}_i^k / \sqrt{\Gamma_i^k}$  approaches zero, we conclude that  $\mathbf{x}_i^{k+1} / \sqrt{\Gamma_i^k}$  approaches zero. Due to the inequality  $\Gamma_i^{k+1} \geq \Gamma_i^k$ ,  $\mathbf{x}_i^{k+1} / \sqrt{\Gamma_i^{k+1}}$  also approaches zero as  $k$  tends to infinity. Since  $\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\| \leq$

$\|\mathbf{x}_i^k\| + \|\bar{\mathbf{x}}_i^k\|$ , the right side of (4.1) approaches zero. Hence, (4.1) implies that  $\mathbf{z}_i^k$  approaches  $\bar{\mathbf{x}}_i^k$  as  $k$  tends to infinity. And since  $\mathbf{z}_i^k$  also approaches  $\mathbf{x}_i^*$  for  $k \in \mathcal{K}$ , we conclude that  $\bar{\mathbf{x}}_i^k$  approaches  $\mathbf{x}_i^*$  as  $k \in \mathcal{K}$  tends to infinity. Let  $\bar{\mathbf{x}}_i^*$  be defined by

$$\bar{\mathbf{x}}_i^* = \arg \min_{\mathbf{u}} \{g(\mathbf{u}) + h(\mathbf{u}) + \langle \mathbf{u}, \mathbf{c}^* \rangle\}.$$

By (4.23) and the fact that  $\bar{\mathbf{x}}_i^k$  approaches  $\bar{\mathbf{x}}_i^*$  as  $k \in \mathcal{K}$  tends to infinity, we conclude that  $\bar{\mathbf{x}}_i^* = \mathbf{x}_i^*$ . In summary, we have

$$\begin{aligned} \lim_{k \in \mathcal{K}} \bar{\mathbf{x}}_i^k &= \mathbf{x}_i^* = \bar{\mathbf{x}}_i^* = \arg \min_{\mathbf{u}} \{g(\mathbf{u}) + h(\mathbf{u}) + \langle \mathbf{u}, \mathbf{c}^* \rangle\}. \\ (4.25) \quad &= \arg \min_{\mathbf{u}} \{f_i(\mathbf{u}) + 0.5\rho\|\mathbf{A}_i\mathbf{u}\|^2 + h_i(\mathbf{u}) + \langle \boldsymbol{\lambda}^* - \rho\mathbf{A}_i\mathbf{x}_i^*, \mathbf{u} \rangle\}. \end{aligned}$$

The first-order optimality conditions for (4.25) are exactly the same as the first-order optimality conditions for (4.20). This shows that (4.20) holds in either Case 1 or Case 2. Hence, (4.21) holds and  $\mathbf{x}^*$  is an optimal solution of (1.1)–(1.2) with associated multiplier  $\boldsymbol{\lambda}^*$ .

Finally, we need to show that the entire sequence converges. If  $\Gamma_i^k$  is uniformly bounded as in Case 1, then by (4.19),  $\mathbf{x}_i^k$  approaches  $\mathbf{x}_i^*$  and  $\|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2/\Gamma_i^k$  approaches zero as  $k$  tends to infinity with  $k \in \mathcal{K}$ . On the other hand, when  $\Gamma_i^k$  tends to infinity as in Case 2, we showed that  $\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2/\Gamma_i^k$  approaches zero. Since  $\bar{\mathbf{x}}_i^k$  for  $k \in \mathcal{K}$  approaches  $\mathbf{x}_i^*$  by (4.25) and  $\Gamma_i^k$  tends to infinity, it follows that  $\|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2/\Gamma_i^k$  approaches zero too. Thus in either Case 1 or Case 2,  $\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2/\Gamma_i^k$  approaches zero as  $k$  tends to infinity with  $k \in \mathcal{K}$ . Letting  $j$  tend to infinity in (4.12) with  $j \in \mathcal{K}$ , it follows that  $E_j$  approaches zero. Moreover, (4.12) implies that along the entire sequence,  $\mathbf{y}_+^k$  approaches  $\mathbf{x}_+^*$  and  $\boldsymbol{\lambda}^k$  approaches  $\boldsymbol{\lambda}^*$ . By (4.13), the entire sequence of iterates  $\mathbf{z}_+^k$  approaches  $\mathbf{x}_+^*$ . Since  $\mathbf{A}\mathbf{z}^k$  approaches  $\mathbf{b}$  (see (4.13)),  $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ , and  $\mathbf{A}_1^\top \mathbf{A}_1$  is invertible, the entire sequence  $\mathbf{z}_1^k$  approaches  $\mathbf{x}_1^*$ . Finally, since  $\mathbf{y}_1^{k+1} = \mathbf{z}_1^k$ , we deduce that the entire  $\mathbf{y}^k$  sequence approaches  $\mathbf{x}^*$ . This completes the proof.  $\square$

**5. Accelerated BOSVS.** The inner loop for the accelerated BOSVS algorithm appears in Algorithm 5.1. As we will see, the inner loop (Step 1) of accelerated BOSVS converges to the minimizer of  $L_i^k$ , exactly as in multistep BOSVS; however, the convergence speed of the multistep BOSVS inner loop is  $O(1/\sqrt{l})$  for the  $\mathbf{z}_i^k$  iterates and  $O(1/l)$  for the objective (see Remark 4.2), while the convergence speed in accelerated BOSVS is  $O(1/l)$  for the  $\mathbf{z}_i^k$  iterates and  $O(1/l^2)$  for the objective, which is optimal for first-order methods applied to general convex, possibly nonsmooth optimization problems.

Two parameter sequences appear in the accelerated BOSVS scheme, the  $\delta^l$  and  $\alpha^l$  sequences. They must be chosen so that the line search condition of Step 1a is satisfied for each value of  $l$ , and the stopping condition of Step 1b is satisfied for  $l$  sufficiently large. If the Lipschitz constant  $\zeta_i$  of  $f_i$  is known, then we could take

$$(5.1) \quad \delta^l = \frac{1}{(1-\sigma)} \frac{2\zeta_i}{l} \quad \text{and} \quad \alpha^l = \frac{2}{l+1} \in (0, 1],$$

in which case, we have

$$\frac{(1-\sigma)\delta^l}{\alpha^l} = \frac{(l+1)\zeta_i}{l} > \zeta_i.$$

This relation along with a Taylor series expansion of  $f_i$  around  $\mathbf{u}_i^{l-1}$  implies that the line search condition in Step 1a of accelerated BOSVS is satisfied for each  $l$ . Moreover,

---

**Initialize:**  $\mathbf{a}_i^0 = \mathbf{u}_i^0 = \mathbf{x}_i^k$  and  $\alpha^1 = 1$ .

**For**  $l = 1, 2, \dots$

1a. Choose  $\delta^l \geq \delta_{\min}$  and when  $l > 1$ , choose  $\alpha^l \in (0, 1)$  such that

$$f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^{l-1} \rangle + \frac{(1-\sigma)\delta^l}{2\alpha^l} \|\mathbf{a}_i^l - \bar{\mathbf{a}}_i^{l-1}\|^2 \geq f_i(\mathbf{a}_i^l),$$

where  $\mathbf{a}_i^l = (1 - \alpha^l)\mathbf{a}_i^{l-1} + \alpha^l\mathbf{u}_i^l$ ,  $\bar{\mathbf{a}}_i^l = (1 - \alpha^l)\mathbf{a}_i^{l-1} + \alpha^l\mathbf{u}_i^{l-1}$ , and

$\mathbf{u}_i^l = \arg \min\{Q(\mathbf{u}) + h_i(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^{n_i}\}$  with

$$Q(\mathbf{u}) = \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u} \rangle + \frac{\delta^l}{2} \|\mathbf{u} - \mathbf{u}_i^{l-1}\|^2 + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u} - \bar{\mathbf{a}}_i^k + \boldsymbol{\lambda}^k / \rho\|^2.$$

1b. If  $\gamma^l = (1/\delta^1) \prod_{j=2}^l (1 - \alpha^j)^{-1} \geq \Gamma_i^{k-1}$ , where  $\gamma^1 = 1/\delta^1$ ,

and  $\|\mathbf{a}_i^l - \mathbf{a}_i^{l-1}\| \leq \psi(e^{k-1})$ , then break.

**Next**

1c. Set  $\mathbf{x}_i^{k+1} = \mathbf{u}_i^l$ ,  $\mathbf{z}_i^k = \mathbf{a}_i^l$ ,  $\Gamma_i^k = \gamma^l$ , and  $r_i^k = (1/\Gamma_i^k) \sum_{j=1}^l \|\mathbf{u}_i^j - \mathbf{u}_i^{j-1}\|^2$ .

---

ALG. 5.1. Inner loop in Step 1 of Algorithm 2.1 for the accelerated BOSVS scheme.

we show (after Lemma 5.2) that with these choices for  $\delta^l$  and  $\alpha^l$ , the stopping condition of Step 1b is also satisfied eventually.

A different, adaptive way to choose the parameters, that does not require knowledge of the Lipschitz constant for  $f_i$ , is the following: Choose  $\delta_0^l \in [\delta_{\min}, \delta_{\max}]$ , where  $0 < \delta_{\min} < \delta_{\max} < \infty$  are safeguard parameters, and set

$$(5.2) \quad \delta^l = \frac{2}{\theta^l + \sqrt{(\theta^l)^2 + 4\theta^l \Lambda^{l-1}}} \quad \text{and} \quad \alpha^l = \frac{1}{1 + \delta^l \Lambda^{l-1}}, \quad \text{where}$$

$$\Lambda^l = \sum_{i=1}^l 1/\delta^i, \quad \Lambda^0 = 0, \quad \text{and} \quad \theta^l = 1/(\delta_0^l \eta^j) \text{ with } \eta > 1.$$

After some algebra, it can be shown that

$$(5.3) \quad \frac{\delta^l}{\alpha^l} = \frac{1}{\theta^l} = \delta_0^l \eta^j.$$

Hence, the ratio  $\delta^l/\alpha^l$  appearing in the line search condition of Step 1a tends to infinity as  $j$  tends to infinity since  $\eta > 1$ . We take  $j \geq 0$  to be the smallest integer for which the line search condition is satisfied. Based on the identity (5.3), the expression  $\delta^l/\alpha^l$  has exactly the same effect as  $\delta_i^k$  in generalized BOSVS. Consequently, it satisfies exactly the same inequality (3.3).

Let us first observe that when  $e^k = 0$ , we have reached a solution of (1.1)–(1.2).

LEMMA 5.1. *If  $e^k = 0$  in the accelerated BOSVS algorithm, then  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{y}^k = \mathbf{z}^k$  solves (1.1)–(1.2) and  $(\mathbf{x}^k, \boldsymbol{\lambda}^k) \in \mathcal{W}^*$ .*

*Proof.* If  $e^k = 0$ , then  $r_i = 0$  for each  $i$ . It follows that

$$(5.4) \quad \mathbf{x}_i^k = \mathbf{u}_i^0 = \mathbf{u}_i^1 = \dots = \mathbf{u}_i^l.$$

By Step 1c,  $\mathbf{u}_i^l = \mathbf{x}_i^{k+1}$ . By the definitions  $\mathbf{a}_i^l = (1 - \alpha^l)\mathbf{a}_i^{l-1} + \alpha^l\mathbf{u}_i^l$  and  $\bar{\mathbf{a}}_i^l = (1 - \alpha^l)\mathbf{a}_i^{l-1} + \alpha^l\mathbf{u}_i^{l-1}$  where  $\mathbf{u}_i^0 = \mathbf{x}_i^k$ , we have  $\mathbf{a}_i^l = \bar{\mathbf{a}}_i^l = \mathbf{x}_i^k$  for each  $l$  due to (5.4).

Again, by Step 1c,  $\mathbf{z}_i^k = \mathbf{x}_i^k$ . Consequently, we have  $\mathbf{x}^{k+1} = \mathbf{x}^k = \mathbf{z}^k$ . Since all three algorithms in this paper share Algorithm 2.1, the remainder of the proof is exactly as in Lemma 3.1.  $\square$

We now establish the following analogue of Lemma 4.2.

LEMMA 5.2. *If the inner loop sequence  $\xi^l := \delta^l \alpha^l \gamma^l$  associated with accelerated BOSVS is nonincreasing as a function of  $l$ , then for each  $i \in [1, m]$ , we have*

$$(5.5) \quad \nu_i \rho \|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\|^2 + \frac{\sigma}{\Gamma_i^k} \sum_{l=1}^{l_i^k} \xi^l \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \leq \frac{\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2}{\Gamma_i^k},$$

where  $l_i^k$  is the terminating value of  $l$  at iteration  $k$ ,  $\bar{\mathbf{x}}_i^k$  is the minimizer of the function  $L_i^k$  defined in (3.8), and  $\nu_i > 0$  is the smallest eigenvalue of  $\mathbf{A}_i^\top \mathbf{A}_i$ .

*Proof.* By the definition  $\mathbf{a}_i^l = (1 - \alpha^l) \mathbf{a}_i^{l-1} + \alpha^l \mathbf{u}_i^l$ , we have

$$\langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^l \rangle = (1 - \alpha^l) \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^{l-1} - \bar{\mathbf{a}}_i^l \rangle + \alpha^l \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \bar{\mathbf{a}}_i^l \rangle.$$

Add to this the identity  $f_i(\bar{\mathbf{a}}_i^l) = (1 - \alpha^l) f_i(\bar{\mathbf{a}}_i^{l-1}) + \alpha^l f_i(\bar{\mathbf{a}}_i^l)$  to obtain

$$\begin{aligned} f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^l \rangle &= \\ (1 - \alpha^l) [f_i(\bar{\mathbf{a}}_i^{l-1}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^{l-1}), \mathbf{a}_i^{l-1} - \bar{\mathbf{a}}_i^{l-1} \rangle] &+ \alpha^l [f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \bar{\mathbf{a}}_i^l \rangle]. \end{aligned}$$

By the convexity of  $f_i$ , it follows that

$$f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^{l-1} - \bar{\mathbf{a}}_i^{l-1} \rangle \leq f_i(\mathbf{a}_i^{l-1}).$$

Hence, we have

$$f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^l \rangle \leq (1 - \alpha^l) f_i(\mathbf{a}_i^{l-1}) + \alpha^l [f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \bar{\mathbf{a}}_i^l \rangle].$$

Adding and subtracting any  $\mathbf{u} \in \mathbb{R}^{n_i}$  in the last term, and then exploiting the convexity of  $f_i$  gives

$$\begin{aligned} f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \bar{\mathbf{a}}_i^l \rangle &= [f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u} - \bar{\mathbf{a}}_i^l \rangle] + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle \\ &\leq f_i(\mathbf{u}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle. \end{aligned}$$

Therefore,

$$(5.6) \quad f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^l \rangle \leq (1 - \alpha^l) f_i(\mathbf{a}_i^{l-1}) + \alpha^l [f_i(\mathbf{u}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle].$$

Now by the line search condition in Step 1a of accelerated BOSVS and then by (5.6), we have

$$\begin{aligned} L_i^k(\mathbf{a}_i^l) &= f_i(\mathbf{a}_i^l) + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{a}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{a}_i^l) \\ &\leq f_i(\bar{\mathbf{a}}_i^l) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{a}_i^l - \bar{\mathbf{a}}_i^l \rangle + \frac{(1 - \sigma) \delta^l}{2 \alpha^l} \|\mathbf{a}_i^l - \bar{\mathbf{a}}_i^l\|^2 \\ &\quad + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{a}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{a}_i^l) \\ &\leq (1 - \alpha^l) f_i(\mathbf{a}_i^{l-1}) + \alpha^l f_i(\mathbf{u}) + \alpha^l \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle + \frac{(1 - \sigma) \delta^l}{2 \alpha^l} \|\mathbf{a}_i^l - \bar{\mathbf{a}}_i^l\|^2 \\ &\quad + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{a}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{a}_i^l). \end{aligned}$$

Next, we utilize the definitions of  $\mathbf{a}_i^l$  and  $\bar{\mathbf{a}}_i^l$  and the convexity of both  $h_i$  and the norm term to obtain

$$\begin{aligned}
L_i^k(\mathbf{a}_i^l) &\leq (1 - \alpha^l)f_i(\mathbf{a}_i^{l-1}) + \alpha^l[f_i(\mathbf{u}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle] + \frac{(1 - \sigma)\delta^l}{2\alpha^l} \|\mathbf{a}_i^l - \bar{\mathbf{a}}_i^l\|^2 \\
&\quad + (1 - \alpha^l) \left( \frac{\rho}{2} \|\mathbf{A}_i \mathbf{a}_i^{l-1} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{a}_i^{l-1}) \right) \\
&\quad + \alpha^l \left( \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{u}_i^l) \right) \\
&= (1 - \alpha^l) \left( f_i(\mathbf{a}_i^{l-1}) + \frac{\rho}{2} \|\mathbf{A}_i \mathbf{a}_i^{l-1} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{a}_i^{l-1}) \right) \\
&\quad + \alpha^l [f_i(\mathbf{u}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle] + \frac{(1 - \sigma)\delta^l \alpha^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \\
&\quad + \alpha^l \left( \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{u}_i^l) \right) \\
&= (1 - \alpha^l) L_i^k(\mathbf{a}_i^{l-1}) + \alpha^l [f_i(\mathbf{u}) + \langle \nabla f_i(\bar{\mathbf{a}}_i^l), \mathbf{u}_i^l - \mathbf{u} \rangle] \\
(5.7) \quad &\quad + \frac{(1 - \sigma)\delta^l \alpha^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 + \alpha^l \left( \frac{\rho}{2} \|\mathbf{A}_i \mathbf{u}_i^l - \mathbf{b}_i^k + \boldsymbol{\lambda}^k / \rho\|^2 + h_i(\mathbf{u}_i^l) \right).
\end{aligned}$$

Since  $h_i$  is convex, we have

$$(5.8) \quad h_i(\mathbf{u}^l) + \langle \mathbf{p}, \mathbf{u} - \mathbf{u}^l \rangle \leq h_i(\mathbf{u})$$

for any  $\mathbf{p} \in \partial h_i(\mathbf{u}^l)$ . The expansion of the quadratic  $Q$  in Step 1a of accelerated BOSVS around  $\mathbf{u}^l$  can be written

$$(5.9) \quad Q(\mathbf{u}^l) + \nabla Q(\mathbf{u}^l)(\mathbf{u} - \mathbf{u}^l) + \frac{1}{2}(\mathbf{u} - \mathbf{u}^l)^\top (\delta^l \mathbf{I} + \rho \mathbf{A}_i^\top \mathbf{A}_i)(\mathbf{u} - \mathbf{u}^l) = Q(\mathbf{u}).$$

Since  $\mathbf{u}^l$  minimizes  $Q + h_i$  in Step 1a, the first-order optimality conditions imply that  $\mathbf{p} + \nabla Q(\mathbf{u}^l) = \mathbf{0}$  for some  $\mathbf{p} \in \partial h_i(\mathbf{u}^l)$ . We choose  $\mathbf{p} = -\nabla Q(\mathbf{u}^l)$ , and then multiply (5.8) and (5.9) by  $\alpha^l$  and add to (5.7) to obtain

$$\begin{aligned}
L_i^k(\mathbf{a}_i^l) &\leq (1 - \alpha^l) L_i^k(\mathbf{a}_i^{l-1}) + \alpha^l L_i^k(\mathbf{u}) + \frac{\delta^l \alpha^l}{2} (\|\mathbf{u} - \mathbf{u}_i^{l-1}\|^2 - \|\mathbf{u} - \mathbf{u}_i^l\|^2) \\
&\quad - \frac{\sigma \delta^l \alpha^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 - \frac{\alpha^l \rho}{2} \|\mathbf{A}_i(\mathbf{u} - \mathbf{u}_i^l)\|^2.
\end{aligned}$$

Hence, for any  $\mathbf{u} \in \mathbb{R}^{n_i}$  we have

$$\begin{aligned}
L_i^k(\mathbf{a}_i^l) - L_i^k(\mathbf{u}) &\leq (1 - \alpha^l)(L_i^k(\mathbf{a}_i^{l-1}) - L_i^k(\mathbf{u})) + \frac{\delta^l \alpha^l}{2} (\|\mathbf{u} - \mathbf{u}_i^{l-1}\|^2 - \|\mathbf{u} - \mathbf{u}_i^l\|^2) \\
(5.10) \quad &\quad - \frac{\sigma \delta^l \alpha^l}{2} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 - \frac{\alpha^l \rho}{2} \|\mathbf{A}_i(\mathbf{u} - \mathbf{u}_i^l)\|^2.
\end{aligned}$$

From the definition of  $\gamma^l$  in accelerated BOSVS, it follows that  $(1 - \alpha^l)\gamma^l = \gamma^{l-1}$  with the convention that  $\gamma^0 = 0$  (since  $\alpha^1 = 1$ ). Hence, for any sequence  $d^l$ ,  $l \geq 0$ , we have

$$(5.11) \quad \sum_{l=1}^j (\gamma^l d^l - (1 - \alpha^l) \gamma^l d^{l-1}) = \sum_{l=1}^j (\gamma^l d^l - \gamma^{l-1} d^{l-1}) = \gamma^j d^j.$$

Suppose that  $d^l \geq 0$  for each  $l$ . By assumption,  $\xi^l = \gamma^l \delta^l \alpha^l$  is nonincreasing; since  $\alpha^1 = 1$  and  $\gamma^1 = 1/\delta^1$ , it follows that  $\xi^1 = 1$ , and we have

$$(5.12) \quad \begin{aligned} \sum_{l=1}^j \xi^l (d^l - d^{l-1}) &= d^1 - d^0 + \sum_{l=2}^j \xi^l (d^l - d^{l-1}) \\ &\geq d^1 - d^0 + \sum_{l=2}^j (\xi^l d^l - \xi^{l-1} d^{l-1}) = \xi^j d^j - d^0. \end{aligned}$$

We now multiply (5.10) by  $\gamma^l$  and sum over  $l$  between 1 and  $l_i^k$ . Exploiting the identity (5.11) with  $d^l = L_i^k(\mathbf{a}_i^l) - L_i^k(\mathbf{u})$  and (5.12) with  $d^l = \|\mathbf{u}_i^l - \mathbf{u}\|^2$ , we obtain

$$(5.13) \quad \begin{aligned} L_i^k(\mathbf{u}) - L_i^k(\mathbf{a}_i^{l_i^k}) &\geq \frac{1}{2\Gamma_i^k} (\xi^{l_i^k} \|\mathbf{u} - \mathbf{u}_i^{l_i^k}\|^2 - \|\mathbf{u} - \mathbf{u}_i^0\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \xi^l \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \\ &\quad + \frac{\rho}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} (\gamma^l \alpha^l) \|\mathbf{A}_i(\mathbf{u} - \mathbf{u}_i^l)\|^2, \end{aligned}$$

where  $\Gamma_i^k$  denotes the final  $\gamma^l$  in accelerated BOSVS.

Next, we multiply the definition  $\mathbf{a}_i^j = (1 - \alpha^j)\mathbf{a}_i^{j-1} + \alpha^j \mathbf{u}_i^j$  by  $\gamma^j$  and sum over  $j$  between 1 and  $l$ . Again, exploiting the identity  $(1 - \alpha^j)\gamma^j = \gamma^{j-1}$  yields

$$(5.14) \quad \mathbf{a}_i^l = \frac{1}{\gamma^l} \sum_{j=1}^l (\gamma^j \alpha^j) \mathbf{u}_i^j.$$

Since  $\alpha^j \gamma^j = \gamma^j - \gamma^{j-1}$ , it follows that

$$(5.15) \quad \gamma^l = \sum_{j=1}^l \alpha^j \gamma^j.$$

Consequently,  $\mathbf{a}_i^l$  is a convex combination of  $\mathbf{u}_i^1$  through  $\mathbf{u}_i^l$ . Since  $\|\mathbf{A}_i(\mathbf{u} - \mathbf{w})\|^2$  is a convex function of  $\mathbf{w}$ , Jensen's inequality yields

$$\frac{1}{\Gamma_i^k} \sum_{l=1}^{l_i^k} (\gamma^l \alpha^l) \|\mathbf{A}_i(\mathbf{u} - \mathbf{u}_i^l)\|^2 \geq \|\mathbf{A}_i(\mathbf{u} - \mathbf{a}_i^{l_i^k})\|^2 = \|\mathbf{A}_i(\mathbf{u} - \mathbf{z}_i^k)\|^2.$$

We apply this inequality to the last term in (5.13) and substitute  $\mathbf{z}_i^k = \mathbf{a}_i^{l_i^k}$ ,  $\mathbf{x}_i^{k+1} = \mathbf{u}_i^{l_i^k}$ , and  $\mathbf{x}_i^k = \mathbf{u}_i^0$  to obtain

$$(5.16) \quad \begin{aligned} L_i^k(\mathbf{u}) - L_i^k(\mathbf{z}_i^k) &\geq \frac{1}{2\Gamma_i^k} (\xi^{l_i^k} \|\mathbf{u} - \mathbf{x}_i^{k+1}\|^2 - \|\mathbf{u} - \mathbf{x}_i^k\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \xi^l \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2 \\ &\quad + \frac{\rho}{2} \|\mathbf{A}_i(\mathbf{u} - \mathbf{z}_i^k)\|^2. \end{aligned}$$

Finally, take  $\mathbf{u} = \bar{\mathbf{x}}_i^k$ . Since the left side of (5.16) is nonpositive for this choice of  $\mathbf{u}$ , the proof is complete.  $\square$

Let us now examine the assumptions and consequences of Lemma 5.2 in the context of the choices (5.1) and (5.2) for the parameters  $\delta^l$  and  $\alpha^l$ . For the choice (5.1) and for  $l \geq 2$ , we have

$$(5.17) \quad \gamma^l = \frac{1}{\delta^1} \prod_{j=2}^l (1 - \alpha^j)^{-1} = \frac{1}{\delta^1} \prod_{j=2}^l \frac{j+1}{j-1} = \frac{1}{\delta^1} \frac{l(l+1)}{2}.$$

Hence,  $\gamma^l$  is  $O(l^2)$ . Since  $\delta^l = \delta^1/l$ , it follows that for  $l \geq 2$ ,

$$\xi^l := \delta^l \alpha^l \gamma^l = \left( \frac{\delta^1}{l} \right) \left( \frac{2}{l+1} \right) \left( \frac{l(l+1)}{2\delta^1} \right) = 1.$$

In the special case  $l = 1$ ,  $\xi^1 = \delta^1/\delta^1 = 1$ . Since the sequence  $\xi^l$  is identically one, it is nonincreasing and the assumption of Lemma 5.2 is satisfied. Since  $\Gamma_i^k$  is the final value for  $\gamma^l$  in Step 1 of accelerated BOSVS, it follows from (5.5) that  $\|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\| = O(1/l_i^k)$ .

For the choice (5.2) and for  $l \geq 2$ , we have  $\Lambda^l = (1/\delta^l) + \Lambda^{l-1}$  and  $\alpha^l = (1/\delta^l)/\Lambda^l$ . It follows that  $1 - \alpha^l = \Lambda^{l-1}/\Lambda^l$  and for  $l \geq 2$ , we have

$$\gamma^l = \frac{1}{\delta^1} \prod_{j=2}^l (1 - \alpha^j)^{-1} = \frac{1}{\delta^1} \prod_{j=2}^l (\Lambda^j / \Lambda^{j-1}) = \frac{1}{\delta^1} \frac{\Lambda^l}{\Lambda^1} = \Lambda^l.$$

Hence,

$$\xi^l := \delta^l \alpha^l \gamma^l = \delta^l \left( \frac{1/\delta^l}{\Lambda^l} \right) \Lambda^l = 1.$$

In the special case  $l = 1$ , we also have  $\xi^1 = 1$ . Again, the sequence  $\xi^l$  is identically one, which satisfies the requirement of Lemma 5.2; consequently, the speed with which  $\mathbf{z}_i^k$  converges to  $\bar{\mathbf{x}}_i^k$  depends on the growth rate of  $\gamma^l$ . By the definition of  $\gamma^l$  in accelerated BOSVS,

$$(5.18) \quad \sqrt{\gamma^l} - \sqrt{\gamma^{l-1}} = \sqrt{\gamma^l} - \sqrt{(1 - \alpha^l)\gamma^l} = \left(1 - \sqrt{1 - \alpha^l}\right) \sqrt{\gamma^l} \geq \frac{\alpha^l \sqrt{\gamma^l}}{2}.$$

Since  $\xi^l := \delta^l \alpha^l \gamma^l = 1$ , it follows from (5.3) that  $(\alpha^l/\theta^l)\alpha^l \gamma^l = (\alpha^l)^2 \gamma^l / \theta^l = 1$ , which implies that

$$(5.19) \quad \alpha^l \sqrt{\gamma^l} = \sqrt{\theta^l}.$$

By (5.18), we have

$$(5.20) \quad \sqrt{\gamma^l} - \sqrt{\gamma^{l-1}} \geq \frac{\sqrt{\theta^l}}{2}.$$

As noted beneath (5.3),  $1/\theta^l$  satisfies the inequality (3.3) for  $\delta_i^k$ , which implies that

$$(5.21) \quad \theta^l \geq \Theta := \frac{1 - \sigma}{\eta \zeta_i + (1 - \sigma)\delta_{\max}}.$$

Hence, (5.20) yields

$$\sqrt{\gamma^l} - \sqrt{\gamma^{l-1}} \geq \frac{\sqrt{\Theta}}{2}.$$

Since  $\gamma^1 = 1/\delta^1 = \theta^1$ , it follows that

$$\sqrt{\gamma^l} \geq \sqrt{\Theta} + \left(\frac{l-1}{2}\right) \sqrt{\Theta} \geq \left(\frac{l}{2}\right) \sqrt{\Theta} \quad \text{or} \quad \gamma^l \geq \left(\frac{l^2}{4}\right) \Theta,$$

which implies that  $\gamma^l = O(l^2)$ . In summary, for either of the choices (5.1) or (5.2), we have  $\xi^l = 1$  for each  $l$ , and  $\|\mathbf{z}_i^k - \bar{\mathbf{x}}_i^k\| = O(1/l_i^k)$ . Moreover, by the inequality (5.16) with  $\mathbf{u} = \bar{\mathbf{x}}_i^k$ , the objective value satisfies  $L_i^k(\mathbf{z}_i^k) - L_i^k(\bar{\mathbf{x}}_i^k) = O(1/(l_i^k)^2)$ .

Although Lemma 5.2 was stated in terms of the terminating iteration  $l_i^k$  of the inner iteration, it applies to any of the inner iterations; that is, for each  $i$  and  $l$ , we have

$$\nu_i \rho \|\mathbf{a}^l - \bar{\mathbf{x}}_i^k\|^2 + \frac{\sigma}{\gamma^l} \sum_{j=1}^l \xi^j \|\mathbf{u}_i^j - \mathbf{u}_i^{j-1}\|^2 \leq \frac{\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2}{\gamma^l}.$$

Whenever  $\gamma^l$  approaches infinity, as it does with the choices (5.1) and (5.2), the right side approach zero and  $\mathbf{a}^l$  converges to  $\bar{\mathbf{x}}_i^k$ . Hence, the stopping conditions in Step 1b of accelerated BOSVS are satisfied for  $l$  sufficiently large when  $e^{k-1} \neq 0$ .

The convergence of accelerated BOSVS, like the other algorithms, relies on a decay property for the iterates, which we now give.

LEMMA 5.3. *If the accelerated BOSVS parameters  $\gamma^l$  tend infinity as  $l$  grows and  $\xi^l := \delta^l \alpha^l \gamma^l = 1$  for each  $l$ , then Lemma 4.4 holds for the accelerated scheme.*

*Proof.* We substitute  $\mathbf{u} = \mathbf{x}_i^*$  and  $\xi^l = 1$  in (5.16) to obtain

$$L_i^k(\mathbf{x}_i^*) - F_i^k(\mathbf{z}_i^k) \geq \frac{1}{2\Gamma_i^k} (\|\mathbf{x}_i^{k+1} - \mathbf{x}_i^*\|^2 - \|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|^2,$$

where  $F_i^k(\mathbf{w}) = L_i^k(\mathbf{w}) + (\rho/2)\|\mathbf{A}_i(\mathbf{w} - \mathbf{x}_i^*)\|^2$ . This is exactly the same as (4.8) in the proof of Lemma 4.4. The remainder of the proof is exactly as in the proof of Lemma 4.4.  $\square$

Using the decay property of Lemmas 4.4 and 5.3, we now obtain the convergence of accelerated BOSVS.

THEOREM 5.4. *Suppose that for the inner loop sequence  $\xi^l := \delta^l \alpha^l \gamma^l$  associated with accelerated BOSVS we have  $\xi^l = 1$  for each  $l$ ,  $\gamma^l$  tends to infinity as  $l$  grows, and there exists a constant  $\kappa > 0$  such that  $\gamma^l(\alpha^l)^2 \geq \kappa$  for all  $l$ . If accelerated BOSVS performs an infinite number of iterations generating iterates  $\mathbf{y}^k$ ,  $\mathbf{z}^k$ , and  $\boldsymbol{\lambda}^k$ , then the sequences  $\mathbf{y}^k$  and  $\mathbf{z}^k$  both approach a common limit  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}^k$  approaches a limit  $\boldsymbol{\lambda}^*$  where  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{W}^*$ .*

*Proof.* The proof is identical to that of Theorem 4.5 through the end of Case 1. For accelerated BOSVS, the fact that  $\mathbf{z}_i^k$  is a convex combination of  $\mathbf{u}_{i,k}^l$  is shown in (5.14)–(5.15). The treatment of accelerated BOSVS first differs from that of multi-step BOSVS in the second paragraph of Case 2 ( $\Gamma_i^k$  tends to  $+\infty$ ) where the multi-step BOSVS stopping condition  $\|\mathbf{u}_i^l - \mathbf{u}_i^{l-1}\|/\sqrt{\gamma^l} \leq \psi(e^{k-1})$ , is used to show that  $\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2/\Gamma_i^k$  approaches zero. Since accelerated BOSVS uses the new stopping condition  $\|\mathbf{a}_i^l - \mathbf{a}_i^{l-1}\| \leq \psi(e^{k-1})$ , a new analysis is needed in Case 2.

By the definition of  $\mathbf{a}^l$ , we have

$$\|\mathbf{a}^l - \mathbf{a}^{l-1}\| = \alpha^l \|\mathbf{u}^l - \mathbf{a}^{l-1}\| \geq \alpha^l (\|\mathbf{u}^l - \mathbf{a}^l\| - \|\mathbf{a}^l - \mathbf{a}^{l-1}\|).$$



If  $\psi_k$  denotes  $\psi(e^{k-1})$  and  $l = l_i^k$  so that  $\mathbf{a}^l$  satisfies the stopping criterion  $\|\mathbf{a}_i^l - \mathbf{a}_i^{l-1}\| \leq \psi_k$ , then

$$\alpha^l \|\mathbf{u}^l - \mathbf{a}^l\| \leq (1 + \alpha^l) \|\mathbf{a}^l - \mathbf{a}^{l-1}\| \leq 2\psi_k \quad \text{or} \quad \|\mathbf{x}_i^{k+1} - \mathbf{z}_i^k\| \leq \frac{2\psi_k}{\alpha^l}$$

since  $\mathbf{u}^l = \mathbf{x}_i^{k+1}$  and  $\mathbf{a}^l = \mathbf{z}_i^k$  when  $l = l_i^k$ . Squaring this, dividing by  $\gamma^l = \Gamma_i^k$ , and utilizing the assumption that  $\gamma^l(\alpha^l)^2 \geq \kappa$  for all  $l$ , we deduce that

$$(5.22) \quad \frac{\|\mathbf{x}_i^{k+1} - \mathbf{z}_i^k\|^2}{\Gamma_i^k} \leq \frac{4\psi_k^2}{\kappa}.$$

Since  $\psi_k$  approach zero by (4.18), it follows that  $\|\mathbf{x}_i^{k+1} - \mathbf{z}_i^k\|^2 / \Gamma_i^k$  approaches zero as  $k$  tends to infinity. Since  $\Gamma_i^k$  is nondecreasing,  $\|\mathbf{x}_i^{k+1} - \mathbf{z}_i^k\|^2 / \Gamma_i^{k+1}$  also approaches zero as  $k$  tends to infinity. Since  $\mathbf{z}_i^k$  is a bounded sequence and  $\Gamma_i^k$  tends to infinity in Case 2, we can replace  $\mathbf{z}_i^k$  by any other bounded sequence and reach the same conclusion. In particular, since the sequence  $\bar{\mathbf{x}}_i^k$  is bounded we conclude that  $\|\mathbf{x}_i^k - \bar{\mathbf{x}}_i^k\|^2 / \Gamma_i^k$  approaches zero as  $k$  tends to infinity, the same conclusion we reached in multistep BOSVS scheme. The rest of the proof is exactly as in Theorem 4.5. This completes the proof.  $\square$

REMARK 5.1. The parameter choices given in both (5.1) and (5.2) satisfy the assumption of Theorem 5.4 that  $\gamma^l(\alpha^l)^2 \geq \kappa > 0$  for some constant  $\kappa$ . In particular, for (5.1), we show in (5.17) that  $\gamma^l = l(l+1)/(2\delta^1)$ . This is combined with the definition of  $\alpha^l$  in (5.1) to obtain

$$\gamma^l(\alpha^l)^2 = \frac{2l}{\delta^1(l+1)} \geq \frac{1}{\delta^1}$$

for  $l \geq 1$ . For the choice (5.2), it follows from (5.19) and (5.21) that

$$\gamma^l(\alpha^l)^2 \geq \Theta := \frac{1 - \sigma}{\eta\zeta_i + (1 - \sigma)\delta_{\max}}.$$

REMARK 5.2. In this paper, we have focused on algorithms based on an inexact minimization of  $L_i^k$  in Step 1 of Algorithm 2.1. In cases where  $f_i$  and  $h_i$  are simple enough that the exact minimizer  $\bar{\mathbf{x}}_i^k$  of  $L_i^k$  can be quickly evaluated, we could simply set  $\mathbf{x}_i^{k+1} = \mathbf{z}_i^k = \bar{\mathbf{x}}_i^k$  and  $\mathbf{r}_i^k = 0$  in Step 1. The analysis of this exact algorithm is very similar to the analysis in Theorems 4.5 and 5.4. In the analysis of the inexact algorithms, a key inequality (4.9) was

$$-\rho \left\langle \sum_{j \leq i} \mathbf{A}_j \mathbf{z}_{e,j}^k + \sum_{j > i} \mathbf{A}_j \mathbf{y}_{e,j}^k + \boldsymbol{\lambda}_e^k / \rho, \mathbf{A}_i \mathbf{z}_{e,i}^k \right\rangle \geq \tau_i^k,$$

where

$$\tau_i^k = \frac{1}{2\Gamma_i^k} (\|\mathbf{x}_{e,i}^{k+1}\|^2 - \|\mathbf{x}_{e,i}^k\|^2) + \frac{\sigma}{2\Gamma_i^k} \sum_{l=1}^{l_i^k} \|\mathbf{u}_{i,k}^l - \mathbf{u}_{i,k}^{l-1}\|^2.$$

For an exact minimizer of  $L_i^k$ , the same inequality can be established but with  $\tau_i^k$  replaced by zero. This follows directly from the first-order optimality conditions for a minimizer of  $L_i^k$  and for a minimizer of (1.1)–(1.2). Since  $\tau_i^k$  disappears, then so do the  $\mathbf{x}_i^k$  and  $\mathbf{u}_{i,k}^l$  terms in Lemma 4.4; consequently, the analysis becomes simpler when the minimizer of  $L_i^k$  is exact.

**6. Numerical Experiments.** In this section, we investigate the performance of the algorithms for an image reconstructed problem that can be formulated as

$$(6.1) \quad \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{F}\mathbf{u} - \mathbf{f}\|^2 + \alpha \|\mathbf{u}\|_{TV} + \beta \|\Psi^T \mathbf{u}\|_1,$$

where  $\mathbf{f}$  is the given image data,  $\mathbf{F}$  is a matrix describing the imaging device,  $\|\cdot\|_{TV}$  is the total variation norm,  $\|\cdot\|_1$  is the  $\ell_1$  norm,  $\Psi$  is a wavelet transform, and  $\alpha > 0$  and  $\beta > 0$  are weights. The first term in the objective is the data fidelity term, while the next two terms are for regularization; they are designed to enhance edges and increase image sparsity. In our experiments,  $\Psi$  is a normalized Haar wavelet with four levels and  $\Psi\Psi^T = I$ . The problem (6.1) is equivalent to

$$(6.2) \quad \min_{(\mathbf{u}, \mathbf{w}, \mathbf{z})} \frac{1}{2} \|\mathbf{F}\mathbf{u} - \mathbf{f}\|^2 + \alpha \|\mathbf{w}\|_{1,2} + \beta \|\mathbf{z}\|_1 \text{ subject to } \mathbf{B}\mathbf{u} = \mathbf{w}, \Psi^T \mathbf{u} = \mathbf{z},$$

where  $\mathbf{B}\mathbf{u} = \nabla \mathbf{u}$  and  $(\nabla \mathbf{u})_i$  is the vector of finite differences in the image along the coordinate directions at the  $i$ -th pixel in the image, while

$$\|\mathbf{w}\|_{1,2} = \sum_{i=1}^N \|(\nabla \mathbf{u})_i\|_2,$$

where  $N$  is the total number of pixels in the image.

The problem (6.2) has the structure appearing in (1.1)–(1.2) with

$$\begin{aligned} f_1(\mathbf{u}) &= 1/2 \|\mathbf{F}\mathbf{u} - \mathbf{f}\|^2, & h_1 &:= 0, \\ f_2 &:= 0, & h_2(\mathbf{w}) &= \|\mathbf{w}\|_{1,2}, \\ f_3 &:= 0, & h_3(\mathbf{z}) &= \|\mathbf{z}\|_1, \\ \mathbf{A}_1 &= \begin{pmatrix} \mathbf{B} \\ \Psi^T \end{pmatrix}, & \mathbf{A}_2 &= \begin{pmatrix} -\mathbf{I} \\ \mathbf{0} \end{pmatrix}, & \mathbf{A}_3 &= \begin{pmatrix} \mathbf{0} \\ -\mathbf{I} \end{pmatrix}, & \text{and } \mathbf{b} &= \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \end{aligned}$$

When solving the test problems using accelerated BOSVS, we use choose  $\alpha^l$  and  $\delta^l$  as in (5.2). Since  $f_2 = f_3 = 0$ , the line search condition holds automatically, and the second and third subproblems are solved in closed form, due to the simple structure of  $h_2$  and  $h_3$ . Only the first subproblem is solved inexactly. At iteration  $k$ , the solution of this subproblem approximates the solution of

$$(6.3) \quad \min_{\mathbf{u}} L_1^k(\mathbf{u}) := \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{b}\|^2 + \frac{\rho}{2} \|\mathbf{B}\mathbf{u} - \mathbf{w}^k + \rho^{-1} \boldsymbol{\lambda}^k\|^2 + \frac{\rho}{2} \|\Psi^T \mathbf{u} - \mathbf{z}^k + \rho^{-1} \boldsymbol{\mu}^k\|^2,$$

where  $\boldsymbol{\lambda}^k$  and  $\boldsymbol{\mu}^k$  are the Lagrange multipliers at iteration  $k$  for the constraints  $\mathbf{B}\mathbf{u} = \mathbf{w}$  and  $\Psi^T \mathbf{u} = \mathbf{z}$  respectively.

The stopping condition for the inner loop of either multistep or accelerated BOSVS required that  $\Gamma_i^k \geq \Gamma_i^{k-1}$ . To improve efficiency, we replaced this condition by

$$l_i^k \geq l_i^{k-1} \quad \text{or} \quad \Gamma_i^k \geq \Gamma_i^{k-1},$$

where  $l_i^k$  is the number of iterations performed by the inner loop for block  $i$  at iteration  $k$ . For all the algorithm, we chose the initial  $\delta_0^l$  in the line search using the BB

approximation, which is given in Step 1a of generalized BOSVS. Moreover, when  $\Gamma_i^k < \Gamma_i^{k-1}$ , we increase  $\delta_{\min,i}$  by setting

$$\delta_{\min,i} := \tau \delta_{\min,i},$$

where  $\tau = 1.1$  in our numerical experiments. When  $\delta_{\min,i}$  is sufficiently large, we have  $\delta_0^l = \delta_{\min,i}$  and the line search condition in the algorithms is satisfied by  $\delta_0^l$ ; that is,  $\delta^l = \delta_0^l = \delta_{\min,i}$ . Consequently, when  $\delta_{\min,i}$  is sufficiently large, we have

$$\Gamma_i^k = \sum_{l=1}^{l_i^k} \frac{1}{\delta^l} = \frac{l_i^k}{\delta_{\min,i}},$$

and the relaxed stopping condition  $l_i^k \geq l_i^{k-1}$  implies that  $\Gamma_i^k \geq \Gamma_i^{k-1}$ , the original stopping condition. Since  $\tau > 1$ , it follows that  $\Gamma_i^k < \Gamma_i^{k-1}$  for only a finite number of iterations, and hence,  $\Gamma_i^k \geq \Gamma_i^{k-1}$  for  $k$  sufficiently large. This ensures the global convergence of the algorithms.

Another improvement to efficiency was achieved by further relaxing the line search criterion. In particular, for the line search in generalized BOSVS (Step 1b), we replaced the right side  $f_i(\mathbf{x}_i^{k+1})$  by  $f_i(\mathbf{x}_i^{k+1}) - \epsilon^k$  where  $\epsilon^k \geq 0$  is a summable sequence. In the line search of multistep BOSVS (Step 1b),  $f_i(\mathbf{u}_i^l)$  was replaced by  $f_i(\mathbf{u}_i^l) - \pi^l$ , where  $\pi^l = \epsilon^k \delta^l \omega^l$  with  $\omega^l$  a summable sequence. In the line search of accelerated BOSVS (Step 1a), we replaced  $f_i(\mathbf{a}_i^l)$  by  $f_i(\mathbf{a}_i^l) - \pi^l$ , where  $\pi^l = \epsilon^k \omega^l / \gamma^l$ . It can be proved that when the line search is relaxed in this way using summable sequences, there is no effect on the global convergence theory; these  $\epsilon^k$  and  $\pi^l$  terms need to be inserted in each inequality in the analysis, but in the end, the steps and the conclusions are unchanged. On the other hand, when the line search is relaxed, it can terminate sooner, and the algorithms can be more efficient. For the numerical experiments, we took  $\epsilon^k = 10/k^{1.1}$ . For multistep BOSVS,  $\omega^l = 1/(\gamma^l)^{1.2}$ , while for accelerated BOSVS,  $\omega^l = 1/(\gamma^l)^{0.6}$ . Since  $\gamma^l = O(1/l)$  for multistep BOSVS and  $\gamma^l = O(1/l^2)$  for accelerated BOSVS, the  $\omega^l$  sequences are summable.

In all the algorithms, we use the following parameters:

$$\delta_{\min} = 10^{-10}, \delta_{\max} = 10^{10}, \alpha = 0.999, \sigma = 10^{-5}, \eta = 3, \text{ and } \tau = 1.1.$$

For the inner loop stopping condition, we took  $\psi(t) = \min\{0.1t, t^{1.1}\}$  in multistep BOSVS, and  $\psi(t) = 0.5t$  in accelerated BOSVS, while in Step 2 of the ADMM template Algorithm 2.1, we took  $\theta_1 = 10^{-6}\sqrt{\rho}$ ,  $\theta_2 = \sqrt{\rho}$ , and  $\theta_3 = 10^{-6}\sqrt{\sigma/(1-\alpha)}$ . For comparison, we provide numerical results based on the algorithm in [23] where we use MATLAB's conjugate gradient routine CGS to solve the subproblem (6.3) almost exactly, stopping when  $\|\nabla L_1^k(\mathbf{u})\| \leq 10^{-6}$ . All the codes were implemented in MATLAB (version R2014a). The following figures show the relative objective error  $(\Phi(\mathbf{u}^k) - \Phi^*)/\Phi^*$  versus CPU time, where  $\Phi^*$  is the optimal function value of (6.1) obtained by applying accelerated BOSVS until the eighth digit of the relative objective value did not change in four consecutive iterations.

The first experiment employs an image deblurring problem from [1]. The original image is the well-known Cameraman image of size  $256 \times 256$  and the observed data  $\mathbf{f}$  in (6.1) is a blurred image obtained by imposing a uniform blur of size  $9 \times 9$  with Gaussian noise and SNR of 40dB. The weights in (6.1) are  $\alpha = 0.005$  and  $\beta = 0.001$ , and the penalty parameter  $\rho = 5 \times 10^{-4}$ . Figure 6.1(a) shows the base-10 logarithm of the relative objective error versus CPU time. In this problem where the subproblems are

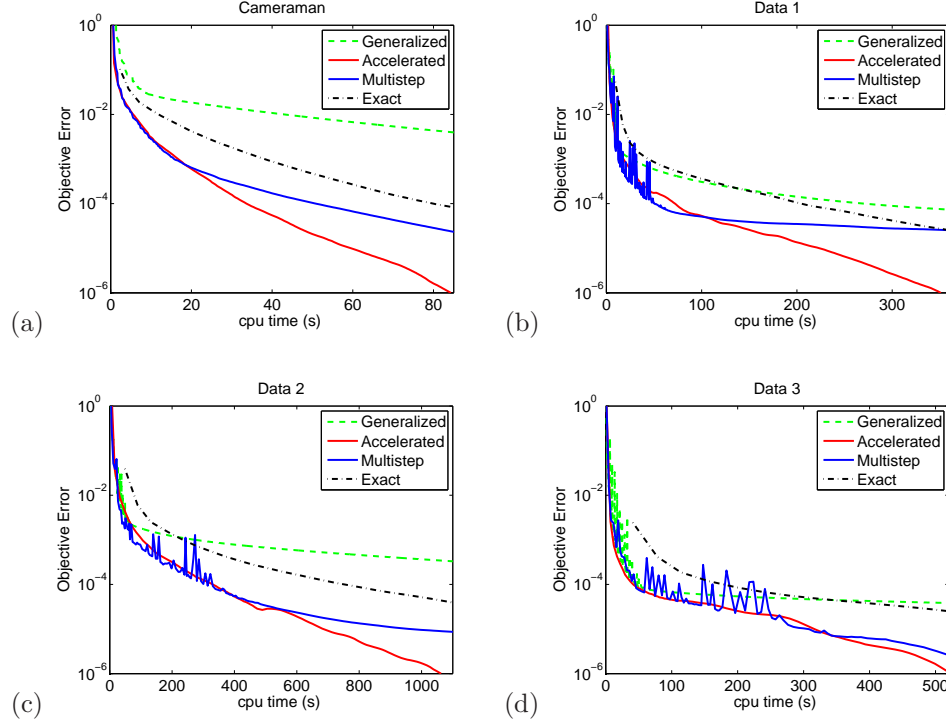


FIG. 6.1. Base-10 logarithm of the relative objective error versus CPU time for the test problems.

relatively easy, generalized BOSVS is significantly slower than the exact, multistep, and accelerated algorithms, while both multistep and accelerated BOSVS were faster than the exact scheme.

The second set of test problems, which arise in partially parallel imaging (PPI), are found in [10]. The observed data, corresponding to 3 different images, are denoted data 1, data 2, and data 3. For these test problems, the weights in (6.1) are  $\alpha = 10^{-5}$  and  $\beta = 10^{-6}$ , and the penalty parameter  $\rho = 10^{-3}$ . The performance of the algorithms is shown in Figure 6.1(b)–(d). These test problems are much more difficult than the first problem since  $\mathbf{F}$  is large, relatively dense, and ill conditioned. In this case, all the inexact algorithms are faster than the exact algorithm initially. The exact algorithm becomes faster than generalized BOSVS when the relative error is around  $10^{-3}$  or  $10^{-4}$ . Accelerated BOSVS is always significantly faster than the exact algorithm.

**7. Conclusion.** Three inexact alternating direction multiplier methods were presented for solving separable convex linearly constrained optimization problems, where the objective function is the sum of smooth and relatively simple nonsmooth terms. The nonsmooth terms could be infinite, so the algorithms and analysis included problems with additional convex constraints. These algorithms all originate from the 2-block variable stepsize BOSVS scheme of [10, 19] which employs indefinite proximal terms and linearized subproblems. The 2-block scheme was generalized to a multi-block scheme using a back substitution process to generate an auxiliary sequence  $\mathbf{y}^k$  that played the role of  $\mathbf{x}^k$  in the original, potentially divergent [5], multiblock ADMM (1.4). The three new methods, called generalized, multistep, and accelerated BOSVS,

correspond to different accuracy levels when solving the ADMM subproblems. Generalized BOSVS employed only one iteration in the subproblems, while multistep and accelerated BOSVS performed multiple iterations until the iteration change was sufficiently small. The multistep and accelerated schemes differed in the rate with which they solved the subproblems. If  $l$  was the number of iterations in the subproblem, then multistep BOSVS had a convergence rate of  $O(1/l)$ , while accelerated BOSVS had a convergence rate of  $O(1/l^2)$ . Global convergence was established for all the methods. Numerical experiments were performed using image reconstruction problems. The accelerated BOSVS algorithm had the best performance when compared with either the other inexact algorithms, or the exact algorithm of [23].

## REFERENCES

- [1] M. AFONSO, J. BIOUCAS-DIAS, AND M. FIGUEIREDO, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356.
- [2] J. BARZILAI AND J. M. BORWEIN, *Two point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.
- [3] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Machine Learning, 3 (2010), pp. 1–122.
- [4] X. CAI, D. HAN, AND X. YUAN, *The direct extension of admm for three-block separable convex minimization models is convergent when one function is strongly convex*, tech. rep., Optimization Online, 2014.
- [5] C. CHEN, B. HE, Y. YE, AND X. YUAN, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Math. Program., 155 (2016), pp. 57–79.
- [6] C. CHEN, M. LI, X. LIU, AND Y. YE, *On the convergence of multi-block alternating direction method of multipliers and block coordinate descent method*, (2015, arXiv:1508.00193).
- [7] C. CHEN, Y. SHEN, AND Y. YOU, *On the convergence analysis of the alternating direction method of multipliers with three blocks*, Abstr. Appl. Anal., 2013 (2013).
- [8] G. CHEN AND M. TEBoulLE, *A proximal-based decomposition method for convex minimization problems*, Math. Programming, 64 (1994), pp. 81–101.
- [9] L. CHEN, D. SUN, AND K. TOH, *An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming*, (2015, arXiv:1506.00741).
- [10] Y. CHEN, W. W. HAGER, M. YASHTINI, X. YE, AND H. ZHANG, *Bregman operator splitting with variable stepsize for total variation image reconstruction*, Comput. Optim. Appl., 54 (2013), pp. 317–342.
- [11] D. DAVIS AND W. YIN, *A three-operator splitting scheme and its optimization applications*, tech. rep., UCLA, Computational and Applied Mathematics Report (CAM), 2015.
- [12] J. ECKSTEIN AND D. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.
- [13] J. ECKSTEIN AND P. J. S. SILVA, *A practical relative error criterion for augmented Lagrangians*, Math. Program., 141 (2013), pp. 319–348.
- [14] J. ECKSTEIN AND W. YAO, *Approximate versions of the alternating direction method of multipliers*, tech. rep., Optimization Online, 2016.
- [15] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.
- [16] R. GLOWINSKI, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, 1984.
- [17] D. GOLDFARB AND S. MA, *Fast multiple-splitting algorithms for convex optimization*, SIAM J. Optim., 22 (2012), pp. 533–556.
- [18] E. GOLSHTEIN AND N. TRETYAKOV, *Modified Lagrangians in convex programming and their generalizations*, in Point-to-Set Maps and Mathematical Programming, P. Huard, ed., vol. 10 of Mathematical Programming Studies, Springer Berlin Heidelberg, 1979, pp. 86–97.
- [19] W. W. HAGER, M. YASHTINI, AND H. ZHANG, *An  $O(1/k)$  convergence rate for the variable stepsize Bregman operator splitting algorithm*, SIAM J. Numer. Anal., (2016, to appear).
- [20] W. W. HAGER AND H. ZHANG, *Projection on a polyhedron that exploits sparsity*, SIAM J.

- Optim., (2015, submitted).
- [21] D. HAN AND X. YUAN, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appl., 155 (2012), pp. 227–238.
  - [22] B. HE, L. LIAO, D. HAN, AND H. YAN, *A new inexact alternating directions method for monotone variational inequalities*, Math. Programming, 92 (2002), pp. 103–118.
  - [23] B. HE, M. TAO, M. XU, AND X. YUAN, *An alternating direction-based contraction method for linearly constrained separable convex programming problems*, Optimization, 62 (2013), pp. 573–596.
  - [24] B. HE, M. TAO, AND X. YUAN, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313–340.
  - [25] B. HE AND X. YUAN, *On the  $\mathcal{O}(1/n)$  convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
  - [26] M. HONG AND Z.-Q. LUO, *On the linear convergence of the alternating direction method of multipliers*, (2012, arXiv:1208.3922).
  - [27] J. LI, G. CHEN, Z. DONG, AND Z. WU, *A fast dual proximal-gradient method for separable convex optimization with linear coupled constraints*, Comput. Optim. Appl., (2016, DOI: 10.1007/s10589-016-9826-0).
  - [28] M. LI, L. LIAO, AND X. YUAN, *Inexact alternating direction methods of multipliers with logarithmic-quadratic proximal regularization*, J. Optim. Theory Appl., 159 (2013), pp. 412–436.
  - [29] M. LI, D. SUN, AND K. C. TOH, *A convergent 3-block semi-proximal ADMM for for convex minimization problems with one strongly convex block*, (2014, arXiv:1410.7933).
  - [30] T. LIN, S. MA, AND S. ZHANG, *On the sublinear convergence rate of multi-block ADMM*, (2014, arXiv:1408.4265).
  - [31] ———, *On the global linear convergence of the ADMM with multiblock variables*, SIAM J. Optim., 25 (2015), pp. 1478–1497.
  - [32] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton Univ. Press, 1970.
  - [33] M. V. SOLODOV AND B. F. SVAITER, *An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions*, Math. Oper. Res., 25 (2000), pp. 214–230.
  - [34] M. TAO AND X. M. YUAN, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57–81.
  - [35] Z. WEN, D. GOLDFARB, AND W. YIN, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Math. Prog. Comput., 2 (2010), pp. 203–230.
  - [36] J. F. YANG AND Y. ZHANG, *Alternating direction algorithms for  $\ell_1$ -problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.